Vibrational, Acoustic and Inertial Sensing on Smart Eyewear for Physiological Monitoring

Angela Cortese[1,*], Federica Mozzini[1,*], Luca Casciano[1,2], Aurelio Teliti[1,2], Matteo Rossi[2], Giacomo Gervasoni[2], Niccolò Antonello[2], Enrico Gianluca Caiani[1]

[1] Politecnico di Milano, Department of Electronics, Information and Bioengineering
Milan, Italy

[2] EssilorLuxottica, EssilorLuxottica Smart Eyewear Lab
Milan, Italy

Abstract

Smart eyewear (sEW) represents a promising platform for physiological monitoring, yet sensor selection remains an open design question. This work characterizes three sensors co-located on the nose pads of a sEW: a vibrational microphone (V2S), an acoustic microphone, and an inertial measurement unit (IMU), evaluated across five tasks (apnea, nasal breathing, eating crunchy and soft food, and drinking) in both quiet and noisy conditions in ten healthy subjects. The V2S achieved the highest accuracy (86.7%) in quiet scenario, while the acoustic microphone collapsed under noise (16.4%), confirming the advantage of contact-based sensing. V2S and IMU fusion yielded the optimal configuration (weighted F1=86.4%). Additionally, V2S, IMU and acoustic microphone were evaluated for heart rate (HR) estimation during apnea: V2S achieved a median error within the 5 bpm frequency resolution limit, outperforming both the IMU, which showed higher variability, and the acoustic microphone, which performed poorly.

CCS Concepts

• **Human-centered computing** → Ubiquitous and mobile computing systems and tools.

Keywords

Smart Eyewear, Wearable Sensing, Physiological Monitoring, Activity Recognition, Sensor Fusion

ACM Reference Format:

Angela Cortese[1,*], Federica Mozzini[1,*], Luca Casciano[1,2], Aurelio Teliti[1,2], Matteo Rossi[2], Giacomo Gervasoni[2], Niccolò Antonello[2], Enrico Gianluca Caiani[1]. 2026. Vibrational, Acoustic and Inertial Sensing on Smart Eyewear for Physiological Monitoring. In *The 24th Annual International Conference on Mobile Systems, Applications and Services (MobiSys Workshop '26), June 21–25, 2026, Cambridge, United Kingdom.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3812836.3814765

1 Introduction

The growing accessibility of wearable devices is reshaping healthcare and lifestyle management, enabling continuous, non-invasive monitoring of physiological and behavioural parameters in everyday settings [1]. Among the available form factors, smart eyewear (sEW) is a particularly promising platform, as glasses are already worn daily by a large fraction of the population [8]. Given their position on the head, sEW are well-suited for monitoring activities involving the nose and mouth, such as breathing, eating and drinking. Unusual breathing patterns and apnea are clinically relevant in sleep-related disorders [20] and as somatic manifestations of psychological stress [3]. Eating and drinking are key behavioural parameters whose automatic and unobtrusive monitoring is essential to overcome the limitations of self-reported dietary assessment [13].

Various sensing modalities have been explored for the detection of these activities across different wearable platforms. Acoustic microphones have been widely used for eating, drinking, and respiratory monitoring, from neck-worn devices [27] to earables [15, 18, 29]. Inertial measurement units (IMUs) have been adopted to capture head and jaw movements associated with eating gestures [5], and prior work has shown that audio and IMU signals carry complementary information for chewing detection [16]. Skin-contact microphones capture bone-conducted signals unavailable to free-field sensing [17]. This principle has supported cardiovascular monitoring: bone-conducted cardiac sounds amplified by the occlusion effect have allowed heart rate (HR) estimation from in-ear microphones [4], while chest-worn contact microphones have demonstrated reliable cardiopulmonary monitoring [4]. Growing attention has recently been given to sEW as a sensing platform. BioGlass [12] demonstrated HR and respiratory rate estimation from IMU and camera signals, while EchoBreath [10] targeted abnormal respiratory symptoms using active and passive acoustic modalities. For dietary monitoring, piezoelectric sensors on the temple [9], IMU and camera combinations [2], integrated acoustic microphones [19], and multimodal fusion [26] have all been explored. However, these works each addressed a single task or a fixed sensor configuration, without systematically comparing the contribution of individual sensing modalities across multiple physiological activities. This limitation is particularly critical in sEW design, where hardware space is constrained, and sensor selection represents a key design decision. Furthermore, vibrational microphones have shown promise for physiological monitoring [11] but, when embedded in sEW, have mostly been used for speech enhancement [6], leaving their applicability to health biomarker sensing uninvestigated.

To address these gaps, our aims were to: *(i)* verify the feasibility of a vibration microphone embedded in sEW for physiological monitoring, including eating and drinking recognition, as well as

[*] These authors contributed equally to this work.

ACM ISBN 979-8-4007-2712-2/2026/06
https://doi.org/10.1145/3812836.3814765

breathing and cardiac activity; *(ii)* perform a comparison of vibration microphone, acoustic microphone, and IMU signals across tasks in both noisy and noise-free conditions; *(iii)* propose an optimized sensor selection strategy, identifying the best sensor configuration for each task.

2 Platform and Data Collection

2.1 Smart Eyewear Platform

To evaluate and compare the performance of vibrational, acoustic, and IMU sensors for physiological monitoring, we utilized a prototypal sEW platform developed by Essilor Luxottica in the Smart Eyewear Lab [1]. Designed primarily as a data acquisition system, the platform facilitates synchronized capture of multimodal signals. In Figure 1, the position of the three primary sensors utilized in this study is highlighted: the LSM6DSV16BX IMU [23] (240 Hz, ±4g / ±1000 dps), the T5838 digital MEMS microphone [25] (48 kHz, 133 dB SPL), and the V2S200D Voice Vibration Sensor (V2S) [24] (48 kHz, ±17.8g).

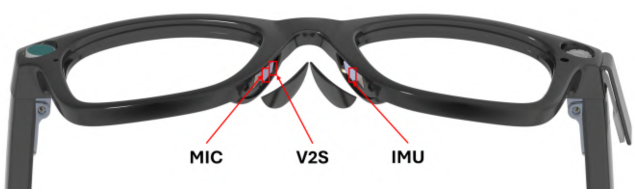

Figure 1: Position of IMU, Acoustic Microphone, and V2S sensors on the sEW platform

Sensors were positioned on the nose pad of the sEW frame, one of the three primary contact points with the wearer, ensuring stable mechanical coupling with the body. This location is proximal to both the nasal airway and the oral cavity, making it suitable for respiratory monitoring and for capturing tissue- and bone-conducted vibrations from mastication and swallowing.

The system is powered by a 200 mAh lithium-polymer battery and orchestrated by an STM32U5 (ARM Cortex-M33) MCU, which manages data acquisition and local storage to a MultiMediaCard (eMMC). Additionally, Bluetooth Low Energy (BLE) connectivity enables remote control and sensor configuration. The acoustic and V2S sensors output their data as pulse density modulation streams at 3.072 MHz, decimated to 48 kHz pulse code modulation via the STM32U5's multi-function digital filter peripheral. The IMU natively outputs 6-axis inertial data (3-axis accelerometer and 3-axis gyroscope) at 240 Hz. Given the multi-modal nature of the platform, strict temporal alignment across all data streams is guaranteed by a global timestamp with 10μs resolution. Timestamped data is routed directly to the eMMC and saved in a raw format for post-processing evaluations.

2.2 Experimental Protocol

Ten subjects (median [25th; 75th percentile], age: 26 [25; 28.5] years, 5F/5M) participated in a controlled data collection inside a sound booth. The study was conducted in accordance with the Declaration of Helsinki (1975, revised 2013), with ethical approval from the Ethics Committee of Politecnico di Milano (opinions n. 85/2025 and n. 86/2025), and all participants provided written informed consent. Each participant was simultaneously equipped with the sEW platform and the eq02+ LifeMonitor (Equivital, Hidalgo Ltd., Cambridge, UK), which provided a 2-lead electrocardiogram (ECG) signal sampled at 256 Hz as a gold standard for cardiac activity. The protocol included five static physiological activities, each performed under silent and noise-exposed conditions, yielding a total of ten experimental segments per subject. In the noise condition, Track 7 of the ICRA Noise corpus [7] was delivered inside the sound booth. The task sequence was as follows: four nasal breaths; breath-holding (apnea); 40 seconds of chewing crunchy food; 40 seconds of chewing soft food; and six sips of water, without volume constraints. Data acquisition was managed through a custom-developed graphical user interface, which enabled the apriori definition of the protocol structure and automatically performed the segmentation of the continuous recordings into task-specific epochs of interest, ensuring consistency and reproducibility across subjects.

[1] https://www.essilorluxottica.com/it/careers/smart-eyewear-lab/

3 Design Rationale

Figure 2 illustrates representative multi-modal recordings of the activities of interest, captured from a single subject wearing the sEW. For each activity, we discuss the underlying physical mechanisms hypothesised to give rise to detectable signals across the three sensing modalities, providing the rationale for the methodology presented in Section 4.

During nasal breathing, turbulent airflow through the nasal passages generates mechanical vibrations that propagate through the nasal soft tissues, which are expected to be detectable by a V2S microphone [15] at the nose pad, while air-conducted pressure fluctuations are captured by the acoustic microphone [10]. Additionally, the IMU is also expected to capture the subtle head displacements associated with the respiratory cycle [12]. Apnea may produce a distinctive absence of respiratory features across all three modalities, as shown in Figure 2(a). A similarly complementary multi-modal signature is expected during drinking: swallowing is accompanied by a discrete vibro-acoustic signature that propagates through the craniofacial soft tissues to the nose pad, allowing the V2S microphone to capture the tissue-conducted component including laryngeal and esophageal movements [29], while the acoustic microphone records airborne emissions from sipping [26]. Head accelerations associated with each sipping gesture are also expected to be detectable by the IMU [26], further complementing the vibro-acoustic information, as illustrated in Figure 2(b). During eating, mastication generates rhythmic jaw movements accompanied by food-texture-dependent acoustic and mechanical emissions propagating through both the air and facial tissues [13, 17]. As shown in Figure 2(c,d), the contrast between crunchy and soft foods is visible: crunchy foods produce high-amplitude, high-frequency transients, whereas soft foods yield lower-intensity, spectrally distinct patterns. The V2S microphone is expected to capture the structure-conducted component of these emissions with reduced susceptibility to environmental noise [2, 16], while the acoustic microphone is expected to record its airborne counterpart [19]; additionally, the periodic head accelerations induced by jaw opening and closing cycles should be detectable by

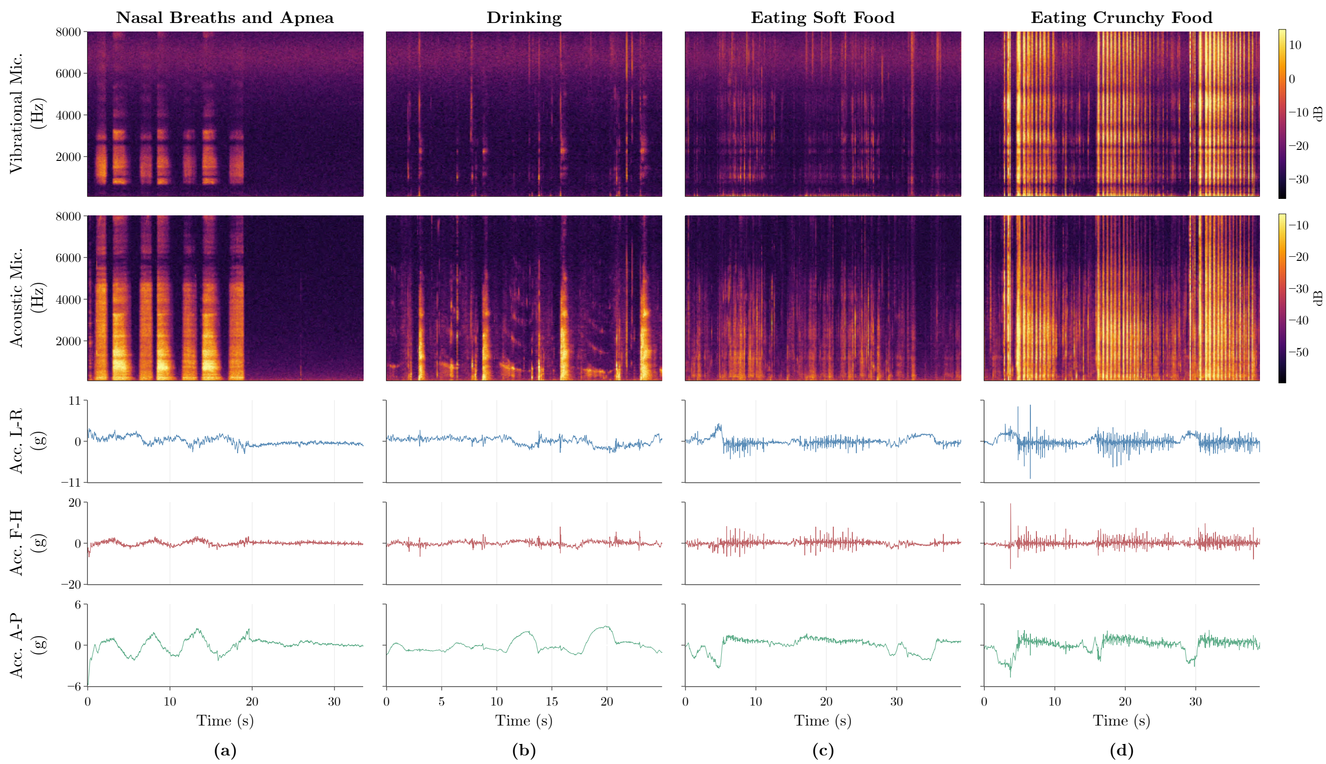

Figure 2: Multi-modal recordings of four tasks performed by a single subject. Each column corresponds to one task: (a) nasal breathing and apnea, (b) drinking, (c) eating soft food, and (d) eating crunchy food. Rows show, from top to bottom: spectrogram of the V2S microphone, spectrogram of the acoustic microphone, and the three axes accelerometer data (L-R: left-right; F-H: front-head; A-P: anterior-posterior).

the IMU [9]. Finally, the nasal region is also particularly favourable for cardiovascular signal acquisition due to the superficial course of the angular artery along the lateral aspect of the nose, where the lack of any significant overlying muscle allows arterial pulsation to remain readily detectable at the skin surface [14]. Given the mechanical nature of arterial pulsation, the IMU and V2S microphone are expected to capture this signal effectively, whereas the acoustic microphone, sensitive to airborne pressure rather than tissue motion, is not expected to provide a reliable signal in this configuration.

4 Methods

4.1 Feature Extraction and Classification

Raw audio signals from V2S and acoustic microphones were downsampled from 48 kHz to 16 kHz, and the first and last 0.5 s of each recording were discarded to remove transient artifacts. Features were extracted using a sliding window of 5 s with 50% overlap. For each microphone signal, 13 Mel-Frequency Cepstral Coefficients (MFCCs) were computed providing a compact representation of the spectral envelope, retaining both the mean and standard deviation (std) across the short-time frames within each 5 s window. For the IMU, features were extracted independently for each of the six axes. Time-domain features include std, inter-quartile range (IQR), skewness, and kurtosis. Frequency-domain features include dominant frequency, spectral entropy, and band power in three bands (0–2 Hz, 2–10 Hz, 10–50 Hz), computed via Fast Fourier Transform (FFT) on each 5 s window. This yields 9 features per axis and 54 features for the full IMU. Four classifiers were evaluated: Random Forest (RF, 100 estimators), Support Vector Machine (RBF kernel, $C = 100$, $\gamma = 0.001$), K-Nearest Neighbors ($k = 5$), and Multilayer Perceptron (hidden layers: 128–64 units), under Leave-One-Subject-Out (LOSO) cross-validation. Noise robustness was assessed via two protocols: *Quiet-to-Noisy* (Q→N) trains on quiet and tests on noisy data, simulating deployment in unseen acoustic conditions; *Mixed-to-Noisy* (QN→N) trains on both quiet and noisy data and tests on noisy, assessing whether noisy training data recovers performance. To evaluate whether combining sensors improves activity recognition, three configurations were compared: IMU alone, V2S microphone alone, and their combination. Fusion was performed at the feature level by concatenating the feature vectors of the individual sensors prior to classification.

4.2 HR Estimation

HR estimation from the V2S, the acoustic microphone, and the F-H acceleration axis [22] of the IMU was performed in the frequency domain and was conducted exclusively during the apnea phase of each trial. Raw signals were bandpass filtered: the V2S and the IMU between 5 and 25 Hz [22], while the acoustic microphone between 20 and 150 Hz [28]. The root mean square (RMS) envelope was then computed over a 50 ms sliding window for all sensors. The RMS signal was segmented to exclude motion artifacts at the onset and offset of each apnea trial. The power spectral density (PSD) was estimated using Welch's method with a fixed window length of 12 s (50% overlap), yielding a consistent frequency resolution of ~5 bpm/bin across all sessions. HR was extracted by identifying the dominant spectral peak within the cardiac band (0.5-2.2 Hz, corresponding to 30−132 bpm). As a ground truth reference, the same spectral procedure was applied to the ECG Lead 1 signal, bandpass filtered between 0.5 and 30 Hz.

5 Results

5.1 Performance in Quiet Conditions

Among the four classifiers evaluated, RF achieved the highest LOSO accuracy across all sensors and was selected for all subsequent experiments. Accuracy is reported as the fraction of correctly classified windows; per-task performance is expressed as weighted F1-score. The V2S microphone achieved the highest overall accuracy (86.7%), followed by the acoustic microphone (82.7%) and the IMU (77.6%). Figure 3 reports the normalized confusion matrices. Apnea was the best-recognized activity across all sensors ($F1 \geq 0.88$), with minimal confusion across all three matrices. Both microphones reliably distinguished *Eat crunchy* from *Eat soft*, while the IMU showed confusion between the two eating tasks (27.8% of *Eat soft* windows predicted as *Eat crunchy*). The IMU performed competitively on *Drink* (F1 = 79.8%), approaching the V2S (F1 = 84.2%). *Breath* was the most confused activity for both the IMU and the acoustic microphone: the IMU misclassified 19.0% of windows as *Eat crunchy* and 11.9% as *Apnea*, while the acoustic microphone spread errors across *Drink* (7.1%), *Eat crunchy* (10.7%) and *Eat soft* (8.3%).

5.2 Noise Robustness

Table 1 summarizes the results of the two noise robustness protocols. One subject was excluded from noisy evaluation due to a recording error in the *Drink* session (N=9 for noisy experiments); the subject was retained in the training set of all remaining folds. Under *Quiet-to-Noisy* (Q⟶N), the acoustic microphone collapsed to near-random performance (16.4%), while the V2S microphone retained strong performance (73.8%). Under *Mixed-to-Noisy* (QN⟶N), adding noisy data to the training set partially recovered the acoustic microphone (56.8%), yet performance remained far below its quiet baseline, confirming that the acoustic signal itself was corrupted by noise rather than the model lacking exposure to noisy conditions. As expected, the IMU was unaffected by acoustic noise by design, showing stable performance across both protocols (72.6% and 73.6%)

5.3 Sensor Fusion

Since the V2S microphone and IMU have been shown to be robust to acoustic noise, sensor fusion experiments used the full dataset (quiet and noisy recordings combined), thus providing a larger and more representative training set. Table 2 reports per-task F1-scores for the three sensor configurations. The combination of V2S and IMU yielded a weighted F1 of 86.4%, outperforming individual sensors (IMU: 76.7%, V2S: 84.8%). The largest gains were on *Drink* and *Eat soft*. The only exception was *Breath*, where V2S alone (87.5%) slightly outperformed fusion (86.8%). *Apnea* remained the easiest task across all configurations (F1: 86.3% − 95.3%).

Table 1: LOSO accuracy under quiet and two noisy protocols. Q⟶N: train on quiet, test on noisy. QN⟶N: train on quiet+noisy, test on noisy.

Sensor	Quiet	Q⟶N	QN⟶N
V2S	**86.7%**	**73.8%**	**81.7%**
MIC	82.7%	*16.4%*	*56.8%*
IMU	77.6%	72.6%	73.6%

Table 2: Per-task F1-score by sensor configuration. Best result per task in bold.

Task	IMU	V2S	V2S+IMU
Apnea	86.3%	93.8%	**95.3%**
Breath	70.0%	**87.5%**	86.8%
Drink	82.7%	83.4%	**88.2%**
Eat crunchy	73.0%	83.9%	**84.0%**
Eat soft	71.6%	78.2%	**80.2%**
Weighted avg	76.7%	84.8%	**86.4%**

*Window counts per class (quiet+noisy): apnea 217, breath 160, drink 235, eat crunchy 281, eat soft 276 (1169 total).

5.4 Sensor Performance for HR Estimation

Apnea duration after temporal cropping ranged from 12.5 to 53.9 s across participants. Given the fixed Welch window length of 12 s, the theoretical frequency resolution of the method is ~5 bpm/bin, which represents the minimum detectable difference in HR estimates. The V2S achieved the best performance (Median Absolute Error, MAE: within one frequency bin, std: 4.8 bpm, max: 15 bpm), with 9 out of 10 sessions within the resolution limit, indicating agreement within one spectral bin. The IMU showed comparable median accuracy (MAE within one frequency bin) but higher variability (std: 16.2 bpm, max: 40 bpm), with 2 sessions showing large errors. The acoustic microphone performed poorly overall (MAE: 22.5 bpm, IQR: 6.3-42.5 bpm, max: 75 bpm).

6 Discussion

The V2S microphone demonstrated a clear ability to discriminate physiological activities, capturing both air turbulence during breathing and vibrations produced by swallowing and chewing through soft tissue transmission. In quiet conditions, the V2S and acoustic microphones achieved comparable performance, both outperforming the IMU. The drinking task was the only exception, where the IMU performed competitively with the V2S, likely because the associated head movements produced a distinctive inertial signal.

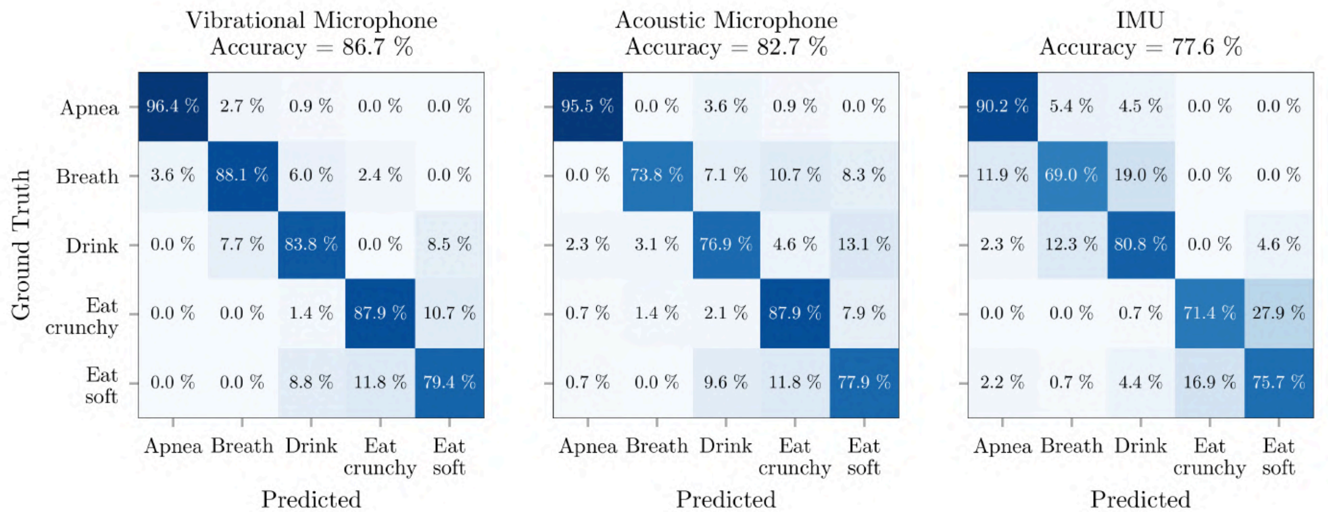

Figure 3: Normalized confusion matrices for V2S, acoustic microphone, and IMU. Random Forest classifier, LOSO cross-validation, quiet conditions (window counts per class: apnea 112, breath 84, drink 130, eat crunchy 140, eat soft 136).

Regarding food discrimination, the V2S successfully distinguished crunchy from soft food consumption, reflecting its sensitivity to mechanical vibrations transmitted through the nose pad. The IMU showed systematic confusion between the two food types, confirming that inertial sensing alone is insufficient to differentiate chewing patterns at this frequency resolution. In noisy conditions, the acoustic microphone collapsed to near-random performance, highlighting a critical limitation for real-world deployment. Even with noisy training data, recovery was only partial, confirming that the acoustic signal itself was corrupted rather than the model lacking adaptation to noise. The V2S retained strong performance across both protocols, confirming the expected robustness of contact-based sensing to airborne noise. The fusion of V2S and IMU yielded the optimal configuration, suggesting the two sensors capture complementary information. It is worth noting that the IMU's robustness under motion artifact remains uninvestigated. For HR estimation, the V2S and IMU confirmed the expected advantage of contact-based sensing, demonstrating that arterial pulsations at the nose pad, despite their small amplitude, are detectable through mechanical coupling, while the acoustic microphone's poor performance was consistent with its insensitivity to tissue-conducted signals (Figure 4).

Beyond performance, the study shows that the V2S and IMU contribute complementary information due to their distinct mechanical coupling pathways and operating frequency ranges. The V2S captures high-frequency tissue- and bone-conducted vibrations (tens–hundreds of Hz), including chewing transients and subtle cardiomechanical components, whereas the IMU measures low-frequency global head and mandibular kinematics (0.1–10 Hz). The combination of these different signal characteristics explains the improvements observed when the two sensors are fused. In contrast, the acoustic microphone, being air-coupled, shows limited coupling to low-amplitude physiological vibrations and is highly susceptible to ambient noise, which accounts for its marked degradation in noisy environments.

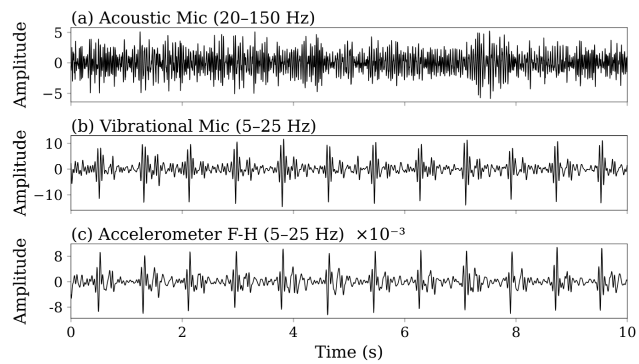

Figure 4: Cardiac activity recordings during apnea: (a) Acoustic microphone, (b) V2S, and (c) Accelerometer F-H axis

7 Conclusion and Future Works

This study presents the first systematic comparison of a V2S microphone, an acoustic microphone, and an IMU mounted on the nose pad of sEW for physiological monitoring. In quiet conditions, both microphones achieve comparable performance; under noise, the acoustic microphone degrades substantially, while the V2S remains robust. The V2S and IMU combination represents the optimal sensor configuration, and when a single sensor is required, the V2S is the most versatile and robust choice. This work also demonstrates for the first time that arterial pulsations at the nose pad are detectable through contact-based sensing, extending prior optical approaches [21] to mechanical sensors.

The present study leaves room for several improvements. The dataset was collected from a limited number of subjects, and larger studies are needed to better assess generalization across diverse populations. Although the acoustic microphone was evaluated under noisy conditions, the robustness of the IMU to motion artifacts, such as those introduced by walking or head movements during daily activities, was not investigated. Future work should incorporate motion conditions across all physiological tasks, enabling a fairer and more complete comparison between sensors. Practical deployment considerations, including power consumption, user comfort, and cost, represent another important direction to explore. Finally, extending HR estimation beyond apnea contexts remains an open and relevant challenge.

Acknowledgments

This work was carried out at the EssilorLuxottica Smart Eyewear Lab, a Joint Research Center between EssilorLuxottica and Politecnico di Milano. FM is attending the PhD program in Biomedical Engineering at the Politecnico di Milano, Cycle XXXIX, with the support of a scholarship financed by the Ministerial Decree no. 117 of 2 March 2023, based on the NRRP (PNRR) Mission 4, Component 2 "From Research to Business" - Investment 3.3 "Introduction of innovative doctorates that respond to the innovation needs of companies and promote the recruitment of researchers from business"

References

[1] Mohan Babu, Ziv Lautman, Xiangping Lin, Milan H.B. Sobota, and Michael P. Snyder. 2024. Wearable Devices: Implications for Precision Medicine and the Future of Health Care. *Annual Review of Medicine* 75 (2024), 401–415. doi:10.1146/annurev-med-052422-020437

[2] Abdelkareem Bedri, Diana Li, Rushil Khurana, Kunal Bhuwalka, and Mayank Goel. 2020. FitByte: Automatic Diet Monitoring in Unconstrained Situations Using Multimodal Sensing on Eyeglasses. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 1–12. doi:10.1145/3313831.3376869

[3] R. Boulding, R. Stacey, R. Niven, and S.J. Fowler. 2016. Dysfunctional breathing: a review of the literature and proposal for classification. *European Respiratory Review* 25, 141 (2016), 287–294.

[4] Kayla-Jade Butkow, Ting Dang, Andrea Ferlini, Dong Ma, and Cecilia Mascolo. 2023. hEARt: Motion-resilient Heart Rate Monitoring with In-ear Microphones. In *2023 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 200–209. doi:10.1109/PERCOM56429.2023.10099317

[5] Garvit Chugh, Indrajeet Ghosh, Sandip Chakraborty, and Suchetana Chakraborty. 2025. BiteSense: Earable-Based Inertial Sensing for Eating Behaviour Assessment. In *2025 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 110–120. doi:10.1109/PerCom64205.2025.00030

[6] Héctor A. Cordourier Maruri, Paulo Lopez-Meyer, Jonathan Huang, Willem Marco Beltman, Lama Nachman, and Hong Lu. 2018. V-Speech: Noise-Robust Speech Capturing Glasses Using Vibration Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4, Article 180 (2018). doi:10.1145/3287058

[7] W. A. Dreschler, H. Verschuure, C. Ludvigsen, and S. Westermann. 2001. ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. *Audiology* 40, 3 (2001), 148–157. PMID: 11465297.

[8] European Council of Optometry and Optics. 2020. Share of Individuals Who Wear Spectacles in Selected European Countries in 2020. Statista. https://www.statista.com/statistics/711514/individuals-who-wear-spectacles-in-selected-european-countries/ Accessed: 19 March 2026.

[9] Muhammad Farooq and Edward Sazonov. 2017. Segmentation and Characterization of Chewing Bouts by Monitoring Temporalis Muscle Using Smart Glasses With Piezoelectric Sensor. *IEEE Journal of Biomedical and Health Informatics* 21, 6 (2017), 1495–1503. doi:10.1109/JBHI.2016.2640142

[10] Kaiyi Guo, Qian Zhang, and Dong Wang. 2025. EchoBreath: Continuous Respiratory Behavior Recognition in the Wild via Acoustic Sensing on Smart Glasses. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, 1–21. doi:10.1145/3706598.3714171

[11] Pranav Gupta, Mohammad J. Moghimi, Yaesuk Jeong, Divya Gupta, Omer T. Inan, and Farrokh Ayazi. 2020. Precision wearable accelerometer contact microphones for longitudinal monitoring of mechano-acoustic cardiopulmonary signals. *npj Digital Medicine* 3, 1 (Feb. 2020), 19. doi:10.1038/s41746-020-0225-7

[12] Javier Hernandez, Yin Li, James M. Rehg, and Rosalind W. Picard. 2014. BioGlass: Physiological Parameter Estimation Using a Head-mounted Wearable Device. In *Proceedings of the 4th International Conference on Wireless Mobile Communication and Healthcare*. IEEE, 55–58. doi:10.4108/icst.mobihealth.2014.257219

[13] H. Hiraguchi, P. Perone, A. Toet, G. Camps, and A.-M. Brouwer. 2023. Technology to Automatically Record Eating Behavior in Real Life: A Systematic Review. *Sensors* 23, 18 (2023), 7757. doi:10.3390/s23187757

[14] A. Letourneau and R. Daniel. 1988. The Superficial Musculoaponeurotic System of the Nose. *Plastic and Reconstructive Surgery* 82, 1 (1988), 48–57.

[15] Yang Liu, Qiang Yang, Kayla-Jade Butkow, Jake Stuchbury-Wass, Dong Ma, and Cecilia Mascolo. 2025. EarMeter: Continuous Respiration Volume Monitoring with Earables. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 4, Article 198 (Dec. 2025), 29 pages. doi:10.1145/3770655

[16] Ramtin Lotfi, George Tzanetakis, Rasit Eskicioglu, and Pourang Irani. 2020. A Comparison Between Audio and IMU Data to Detect Chewing Events Based on an Earable Device. In *Proceedings of the 11th Augmented Human International Conference (AH '20)*. ACM, 1–8. doi:10.1145/3396339.3396362

[17] Akihiro Nakamura, Takato Saito, Daizo Ikeda, Ken Ohta, Hiroshi Mineno, and Masafumi Nishimura. 2021. Automatic Detection of Chewing and Swallowing. *Sensors* 21, 10 (2021), 3378. doi:10.3390/s21103378

[18] V. Papapanagiotou, S. Ganotakis, and A. Delopoulos. 2021. Bite-Weight Estimation Using Commercial Ear Buds. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, Mexico, 7182–7185. doi:10.1109/EMBC46164.2021.9630500

[19] Vasileios Papapanagiotou, Anastasia Liapi, and Anastasios Delopoulos. 2022. Chewing Detection from Commercial Smart-glasses. In *Proceedings of the 7th International Workshop on Multimedia Assisted Dietary Management (MADiMa '22)*. ACM, 11–16. doi:10.1145/3552484.3555746

[20] Dimitrios A. Pappas, Jon T. Giles, Geoffrey Connors, Noah Lechtzin, Joan M. Bathon, and Sonye K. Danoff. 2010. Respiratory Symptoms and Disease Characteristics as Predictors of Pulmonary Function Abnormalities in Patients with Rheumatoid Arthritis: An Observational Cohort Study. *Arthritis Research & Therapy* 12 (2010), 1–11.

[21] A. Scandelli, I. Crupi, P. Bartoli, A. Giudici, A. De Vecchi, and F. Villa. 2024. Study of PPG Sensor Positioning on Smart Eyewear for Biosensing. In *Proceedings of the 2024 IEEE Sensors Applications Symposium (SAS)*. IEEE, Naples, Italy, 1–6. doi:10.1109/SAS60918.2024.10636483

[22] S. Solbiati et al. 2024. Comparison of ECG-Free Algorithms for Heart Rate Computation From Head-BCG Signals Obtained With Smart Eyewear. In *2024 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*. 213–218. doi:10.1109/MetroXRAINE62247.2024.10795866

[23] STMicroelectronics 2024. *LSM6DSV16BX: 6-axis IMU with embedded sensor fusion, AI, Qvar and hearable features for TWS*. STMicroelectronics. https://www.st.com/resource/en/datasheet/lsm6dsv16bx.pdf Rev 4.

[24] Syntiant 2026. *V2S200D: multimode digital vibration sensor*. Syntiant. https://static1.squarespace.com/static/6488b0b8150a045d2d112999/t/697230321a986a77fb11c224/1769091122500/Syntiant+V2S200D+Datasheet+Rev+C.pdf Rev C.

[25] TDK 2025. *T5838: Bottom Port PDM Digital Output Multi-Mode Microphone*. TDK. https://invensense.tdk.com/wp-content/uploads/2025/10/DS-000383-T5838-Datasheet-v1.2.pdf Rev 1.2.

[26] Ruiqi Wang, Gerardo Gonzalez, Wei Liu, Chao Wang, and Tsung-Sheng Tseng. 2025. DrinkMR: Unobtrusive Drinking Monitoring Using Off-The-Shelf Smart Glasses. In *2025 IEEE 22nd International Conference on Mobile Ad-Hoc and Smart Systems (MASS)*. IEEE, 93–101. doi:10.1109/MASS66014.2025.00026

[27] Koji Yatani and Khai N. Truong. 2012. BodyScope: A Wearable Acoustic Sensor for Activity Recognition. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp)*. ACM, 341–350. doi:10.1145/2370216.2370269

[28] Jae-Young Yoo et al. 2023. Wireless broadband acousto-mechanical sensing system for continuous physiological monitoring. *Nature Medicine* 29, 12 (2023), 3137–3148. doi:10.1038/s41591-023-02637-5

[29] Shuyu Zhang, Yunfan Liu, and Mahanth Gowda. 2022. Let's Grab a Drink: Teacher-Student Learning for Fluid Intake Monitoring using Smart Earphones. In *2022 IEEE/ACM Seventh International Conference on Internet-of-Things Design and Implementation (IoTDI)*. IEEE, 55–66. doi:10.1109/IoTDI54339.2022.00014