A Novel HMD-Mounted Contactless Proxi-rPPG Sensor for Emotion Recognition in Free-Moving VR Environments

Heesook Shin†
Content Research Laboratory
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
hsshin8@etri.re.kr

Yongho Lee
Content Research Laboratory
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
jason0720@etri.re.kr

Seoyeon Lim
Department of Information Technology Engineering
Sookmyung Women's University
Seoul, Republic of Korea
imseoyeon0218@sookmyung.ac.kr

Seung Hyun Song
Department of Electrical Engineering
Sookmyung Women's University
Seoul, Republic of Korea
shsong.ee@sookmyung.ac.kr

Suh-Yeon Dong
Division of Artificial Intelligence Engineering
Sookmyung Women's University
Seoul, Republic of Korea
sydong@sookmyung.ac.kr

Youn-Hee Gil
Content Research Laboratory
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea
yhgil@etri.re.kr

Abstract

In this paper, we propose Proxi-rPPG—a contactless remote photoplethysmography (rPPG) sensor mounted on a head-mounted display (HMD)—to reliably obtain heart rate data in a virtual reality (VR) environment where users can move freely. We also develop and evaluate an emotion-recognition model using the Proxi-rPPG approach. Specifically, we attach a small camera module to the outside of an HMD to capture the user's cheek–ear region at close range, then extract heart rate and heart rate variability via an rPPG algorithm. While traditional reflective PPG requires precise skin contact—and wrist-worn sensors can be highly susceptible to noise from arm movements—conventional rPPG methods also face limitations imposed by user motion. Proxi-rPPG addresses these drawbacks and demonstrates its potential to quantify emotional changes arising from VR content experiences. In a user study with 20 participants, we compared waveform data from the Proxi-rPPG sensor to that of a reflective wrist-PPG device. We found that the Proxi-rPPG approach yielded a lower noise ratio than the wrist sensor in dynamic VR conditions. It also achieved 65 % accuracy in a four-quadrant emotion classification task (High/Low Arousal × Positive/Negative Valence) using a machine-learning model. These findings suggest that Proxi-rPPG can be effectively applied to various domains, including emotion monitoring in VR.

CCS Concepts

• Ubiquitous and mobile computing design and evaluation methods • HCI design and evaluation methods

†Corresponding author

UbiComp Companion '25, October 12–16, 2025, Espoo, Finland

ACM ISBN 979-8-4007-1477-1/2025/10.
https://doi.org/10.1145/3714394.3754394

Keywords

Proxi-rPPG sensor, Remote PPG, Virtual reality, Emotion recognition

ACM Reference format:

Heesook Shin, Yongho Lee, Seoyeon Lim, Seung Hyun Song, Suh-Yeon Dong and Youn-Hee Gil. 2025. A Novel HMD-Mounted Contactless Proxi-rPPG Sensor for Emotion Recognition in Free-Moving VR Environments. In Proceedings of the 2025 ACM International Symposium on Wearable Computers (ISWC'25), October 12-16, 2025, Espoo, Finland. ACM, New York, NY, USA, XX pages. https://doi.org/10.1145/3714394.3754394

1 Introduction

Virtual reality (VR) technology is widely applied in gaming, education, and rehabilitation, offering high immersion and free user interaction [1, 2, 3]. Real-time understanding of a user's emotional state—such as focus, stress, or engagement—could enable automatic adjustments in content difficulty or help improve affective expression and empathy in training scenarios.

Emotion recognition requires comprehensive analysis of various signals including facial expressions, vocal cues, physiological data, and motion. Facial-expression–based emotion recognition remains significant, yet head-mounted displays (HMDs) occlude the upper face, complicating detection [4]. Techniques to address this challenge in VR environments include robust deep-learning models designed for partially occluded faces [5], real-time lower-face analysis in immersive VR [6], and advanced image-processing approaches to recover occluded facial features [7]. Meanwhile, other physiological signals (electroencephalography (EEG), photoplethysmography (PPG), electromyography (EMG)) have also been proposed [8, 9], and some work combines multiple modalities to enhance accuracy [10, 11]. Among these signals, heart rate (HR) and heart rate variability (HRV) from PPG effectively track VR users' emotional states.

Some studies integrate PPG sensors directly into HMDs (e.g., PhysioHMD [12], EmteqVR [13]) by placing reflective modules in contact with facial skin, though reports indicate reduced comfort and a risk of detachment. Wrist-worn PPG sensors are simpler to use but produce significant motion noise when arms move and require additional gear [8]. Remote PPG (rPPG) estimates blood volume pulses from camera images (face, wrist, etc.) without direct skin contact [14]. While early methods worked well only in stable lighting and minimal motion, subsequent algorithms like POS (Plane-Orthogonal to Skin), CHROM (chrominance-based), and PCA (Principal Component Analysis)-based corrections [15, 16] have broadened rPPG's applicability. Still, user head rotations or movements outside the camera's field of view present challenges in free-moving VR contexts.

To address these limitations, we introduce Proxi-rPPG—a contactless rPPG sensor attached to the exterior of an HMD. Specifically, a small camera module on its side or front captures the user's cheek, ear, and temple areas at close range to extract PPG signals without skin contact. Proxi-rPPG offers greater comfort than reflective PPG, reduces noise from large head movements, and overcomes the limitations of conventional rPPG, which requires a fixed camera and limits user mobility. It can also be extended to glasses, AR devices, and other wearable platforms.

Accordingly, this paper presents (1) the design and prototype of Proxi-rPPG sensor for VR, (2) a comparative performance study against a wrist-mounted reflective PPG sensor, and (3) an emotion recognition model using HR data, with potential applications in affective and social interaction training.

2 Proxi-rPPG Sensor Implementation

2.1 HW Specification

A simple prototype of the Proxi-rPPG sensor was constructed using a small RGB camera (Raspberry Pi Camera Module v2 NoIR, 1920×1080 @24fps) and a wireless-enabled microprocessor (Raspberry Pi 4 Model B). It was magnetically attached to the side light-shielding section of a Meta Quest Pro HMD, as illustrated in Figure 1.

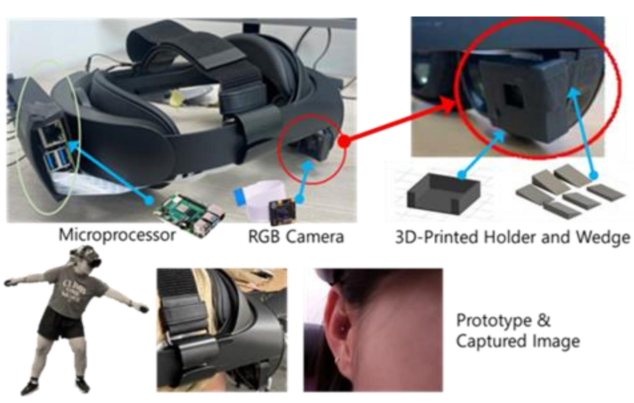


Figure 1: Prototype Proxi-rPPG sensor magnetically mounted on a Meta Quest Pro HMD

A custom camera mount and an adjustable wedge were designed to match each user's facial structure, ensuring that the camera module could capture the cheek–ear region at a diagonal angle from about 3–5 cm away. Additionally, to account for the mixture of physiological changes induced by both emotional responses (e.g., fear, excitement) and physical activity (movement) during VR content, we incorporated a contactless temperature sensor. Specifically, a thermopile sensor that absorbs mid-IR radiation emitted from the skin was chosen, and its data were transmitted over the same wireless channel as the image signals. Although this study does not present detailed analysis of the temperature data, we plan to utilize it in future research aimed at mitigating motion-related artifacts in PPG signals.

2.2 Extraction of Heart Rate Using Proxi-rPPG

Based on prior studies, we identified regions that effectively reflect blood-flow variations—namely the cheek, ear (helix and lobe), and temple. In particular, we drew on the facial surface orientation method proposed in [17] and the region-based heart rate measurement approach in [18], ultimately setting the region of interest (ROI) to the cheek–ear area. With the HMD on, we used a camera preview to ensure that the ear and cheek remained centered on the display. Through a pilot test (N = 5), we set the camera's field of view (FoV) to approximately 45 ± 5° so that the ear and cheek would remain within the frame. We targeted areas not occluded by the HMD as the ROI. Pixel coordinates of x: 0 – 450 and y: 300 – 700 were determined to be the most robust to movement and lighting variations.

The collected video (30fps) was processed by a standard rPPG pipeline, applying a green-channel analysis and a Plane-Orthogonal to Skin (POS) algorithm [15, 16] to correct for lighting and skin reflectivity. Afterward, we applied a 0.7–4Hz bandpass filter and peak detection to calculate HR and HRV. We anticipate that further adopting a deep learning–based HR estimation method in place of—or in combination with—this pipeline could enhance robustness to lighting changes and motion while improving overall accuracy and efficiency [19, 20, 27]. Under static user conditions, the Proxi-rPPG prototype demonstrated high reliability, exhibiting a mean error of 0.052 bpm (Reference HR: 87.667 bpm, Proxi-rPPG HR: 87.791 bpm, N=5) when compared with a reference device (Polar Verity Sense).

3 Experimental Procedure

3.1 Experimental Setup

To validate the performance of the Proxi-rPPG sensor and explore its potential for emotion recognition, we conducted an experiment in which participants experienced immersive VR content. We compared the Proxi-rPPG sensor's performance with that of a wrist-worn PPG device (in smartwatch form). Twenty adults (10 males, 10 females; mean age 23.2±2.33 years)—who reported no severe VR sickness and no visual or auditory impairments—were recruited through an open call, provided they had used a VR device at least once prior to the study. For emotion elicitation, we consulted existing research on VR content known to induce emotional responses, then performed a pilot test (N=5) to finalize four representative VR titles spanning four different genres.

Table 1: The four types of immersive, interactive VR content

Emotion Classification	Representative Content and Description	Representative Image
High Arousal + Positive Valence (HAPV)	(Action/Sports): "CARVE Snowboarding" [23], A VR title where you descend a mountain slope on a snowboard	
High Arousal + Negative Valence (HANV)	(Horror/Thrill): "Affected The Manor" [24], A VR title involving exploration of a haunted mansion	
Low Arousal + Positive Valence (LAPV)	(Meditation/Relaxation): "Fujii" [25], A VR title focusing on watering plants, gently interacting with living creatures, and restoring an ecosystem	
Low Arousal + Negative Valence (LANV)	(Somber/Monotone): "Blair Witch: Oculus Quest Edition" [26], A VR title where you play a police officer searching for a missing child	

A 9-point Self-Assessment Manikin (SAM) scale was used to collect participants' subjective emotional responses. SAM enables intuitive assessment of Valence (1 = extremely negative, 9 = extremely positive) and Arousal (1 = very calm, 9 = highly excited). Each participant viewed a SAM figure and marked a numeric rating immediately after completing each VR content segment. This study specifically collected Valence—representing the continuum from positivity to negativity—and Arousal, measuring the level of emotional activation from High to Low.

3.2 Experimental Protocol

The experiment employed a Latin Square design so that participants experienced the four VR contents (HAPV, HANV, LAPV, LANV) in different orders. In the Preparation phase (20 minutes), participants received an overview of the study's objectives, procedures, and potential risks, as well as instructions on how to complete the SAM questionnaire. They then donned the HMD with the Proxi-rPPG sensor prototype attached, while a Polar Verity Sense devices were worn on both wrists and ankles. Four Polar units logged tri-axial accelerometer (ACC) and gyroscope (GYRO) data to quantify upper- and lower-limb motion, and the right-wrist unit simultaneously recorded reflective contact PPG, providing the reference heart rate (reference PPG) signal. Participants also used the official Meta tutorial content [21] to familiarize themselves with VR controller operations.

Each Trial phase (20 minutes) comprised Fixation, VR content, and Recovery sessions for each of the four contents. At the start of each trial, participants performed a Fixation session, aligning their position with the VR content's starting viewpoint. They then experienced one of the four VR contents for about 10 minutes (VR content session). Afterward, a 10-minute Recovery session was provided to allow participants to rest and evaluate the emotional impact of the prior VR exposure, during which they completed the SAM questionnaire to provide subjective emotional ratings.

Figure 2: Overview of the Experimental Protocol

For data collection, the Raspberry Pi on the HMD side tagged each video frame with a normalized timestamp at one-second intervals via Wi-Fi, while the Polar device assigned timestamps to its Bluetooth Low Energy (BLE) packets, enabling alignment to a common time axis. After the experiment ended, the PPG signals captured by the camera, along with calculated HR data, motion data (ACC, GYRO), temperature data (thermopile sensor data), and SAM (1–9) labels were merged based on matching timestamps.

4 Experimental Results and Analysis

4.1 Evaluation of the Proxi-rPPG

After each VR content segment, participants entered a recovery period and performed a single right-arm swing to provoke motion artifacts. Analysis of both the entire one-minute signal and a magnified 10-s window showed that the reflective wrist-worn PPG waveform oscillated in step with peaks from the ACC, whereas the contactless Proxi-rPPG trace remained comparatively stable over the same 10-s (240-frame) sequence. Quantitatively, the coefficient of variation ($CV = \sigma/\mu \times 100\,\%$) of the wrist PPG was ~30 %, while that of the Proxi-rPPG was only ~10 %, confirming a three-fold reduction in motion-induced noise. A temporal offset between the ACC and reference PPG peaks—despite their common hardware origin—likely stems from sampling-rate mismatches or transient contact and muscle artifacts during movement [22].

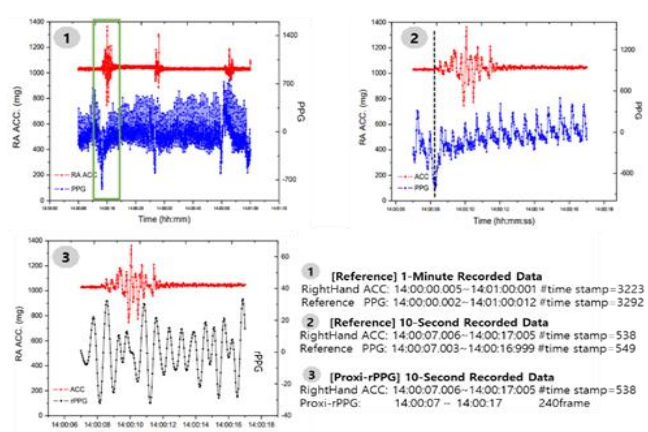
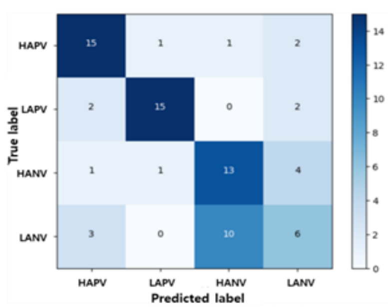


Figure 3: Comparison of Reference PPG and Proxi-rPPG

4.2 Emotion Recognition Model

Next, we examined whether an emotion recognition model could be constructed using data from the Proxi-rPPG sensor. We extracted both time-domain and frequency-domain features from the collected waveform signals, including RMSSD (Root Mean Square of the Successive Differences), SDNN (Standard Deviation of NN intervals), LF (Low Frequency Band), HF (High Frequency Band), LF/HF, mean, and standard deviation. These were employed as input features in machine-learning classification models—Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (KNN), and Logistic Regression (LR)—under both Early Fusion and Late Fusion approaches. In this paper, Early Fusion combines all features into one classifier, whereas Late Fusion first classifies each modality independently and then fuses the outputs.

In the Early Fusion results, the following accuracies were obtained: DT (56.58 %), SVM (59.42 %), RF (64.75 %), KNN (54.17 %), and LR (55.33 %). The RF model achieved the highest performance (64.75 %) when using four signals: HR, temperature, ACC, and GYRO. Figure 4 summarizes the RF model's test performance via 5-fold cross-validation, indicating that HR data derived from the sensor prototype can be effectively used to classify the four emotional states.



	precision	recall	f1-score
HAPV	0.71	0.79	0.75
LAPV	0.88	0.79	0.83
HANV	0.54	0.68	0.60
LANV	0.43	0.32	0.36

Figure 4: Confusion Matrix and Classification Results of the RF Model Using HR+Temp+ACC+GYRO Features

From the confusion matrix and the classification results, one can see that LANV is occasionally misclassified as HANV, which is consistent with the questionnaire findings. Although participants' SAM survey responses generally aligned with each content's genre, a relatively large standard deviation was found for HANV in the Valence dimension as shown in Figure 5.

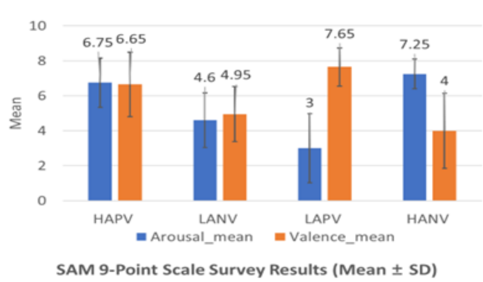


Figure 5: Participants' SAM Ratings Summary

Moreover, a detailed breakdown shows that eight participants labeled the LANV content as HAPV, LAPV, or HANV, representing the highest rate of unintended responses among the four contents. We presume this stems from the small pilot test sample (N=5), which may have been insufficient to fully validate the negative affect induction intended by LANV. Future research should include a larger-scale verification of LANV content or compare additional somber/negative VR materials.

Under the Late Fusion approach, we classified the pulse-wave signals (Proxi-rPPG) with SVM and the GYRO data with RF, achieving a maximum accuracy of 63.15 %. This again indicates that the waveforms from the Proxi-rPPG sensor can serve as valid features for an emotion-recognition model. As observed in Early Fusion, LANV was also frequently misclassified as HANV in the Late Fusion results.

5 Conclusion and Future Work

This study is significant in that it proposes Proxi-rPPG, a contactless, proximity-based rPPG sensor, for use in VR environments where users can move freely. Specifically, we integrated a small camera module onto an HMD to capture signals from the user's cheek–ear region at close range, ensuring stable heart rate extraction despite head movements. Through a 20-participant evaluation, we showed that this approach addresses the signal distortion typically associated with contact-based PPG systems, while demonstrating the feasibility of biometric-driven emotion recognition without requiring additional worn devices. Moreover, we achieved 65 % accuracy in a four-quadrant emotion classification task, confirming that a headset-only, contactless Proxi-rPPG sensor can track affective changes even while users move.

To raise this figure to a level suitable for real-world deployment, we will pursue three complementary upgrades. (i) Hardware: redesign the mount so that the region of interest shifts from the cheek–ear region to the auricular helix/lobe, fully utilize the existing thermopile module, and improve optical shielding, thereby increasing pulsatility and mitigating artifacts caused by hair, accessories, and lighting. (ii) Stimuli: replace the ambiguous LANV title with VR scenarios whose valence–arousal labels are pre-validated in a larger pilot, reducing label noise. (iii) Algorithms: substitute the current Random Forest pipeline with a multimodal Transformer that fuses HR, temperature, ACC, and GYRO streams; this architecture has delivered 5–10 percentage-point gains on comparable affective-computing benchmarks [27]. These advances are expected to push accuracy beyond 65 % and enable robust, real-time emotion monitoring in applications ranging from VR-based training and digital therapeutics to remote collaboration.

Acknowledgements

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government [25ZC1200, Research on hyper-realistic interaction technology for five senses and emotional experience].

References

[1] Federica Pallavicini, Alessandro Pepe, and Maria Eleonora Minissi. 2019. Gaming in virtual reality: What changes in terms of usability, emotional response and sense of presence compared to non-immersive video games?. Simulation & Gaming, Vol. 50, No. 2, 136-159. https://doi.org/10.1177/1046878119831420

[2] Rukshani Somarathna, Tomasz Bednarz and Gelareh Mohammadi. 2023. Virtual Reality for Emotion Elicitation – A Review. IEEE Transactions on Affective Computing, Vol. 14, No. 4. https://doi.org/10.1109/TAFFC.2022.3181053

[3] Heesook Shin et al. 2024. Effect of Virtual Intervention Technology in Virtual Vocational Training for People with Intellectual Disabilities: Connecting Instructor in the Real World and Trainee in the Virtual World. International Journal of Human-Computer Interaction, Vol. 40, No. 3, 624-639. https://doi.org/10.1080/10447318.2022.2121560

[4] Ifigeneia Mavridou et al. 2017. FACETEQ interface demo for emotion expression in VR. IEEE Virtual Reality, Los Angeles, CA, 441-442. https://doi.org/10.1109/VR.2017.7892369

[5] Palanichamy Naveen. 2024. Occlusion-aware Facial Expression Recognition: A Deep Learning Approach. Multimedia Tools and Applications, Vol. 83, 32895–32921. https://doi.org/10.1007/s11042-023-17013-1

[6] Xinrun Chen and Hengxin Chen. 2023. Emotion Recognition Using Facial Expressions in an Immersive Virtual Reality Application. Virtual Reality, Vol. 27, 1717–1732. https://doi.org/10.1007/s10055-022-00720-9

[7] Qingzhen Gong, Xuefang Liu and Yongqiang Ma. 2024. Real-Time Facial Expression Recognition Based on Image Processing in VR. Cognitive Neurodynamics. https://doi.org/10.1007/s44196-024-00729-9

[8] Zhilin Zhang, Zhouyue Pi and Benyuan Liu. 2015. Troika: A General Framework for Heart Rate Monitoring Using Wrist-Type Photoplethysmography Signals During Intensive Physical Exercise. IEEE Transactions on Biomedical Engineering, Vol. 62, No. 2, 522–531. https://doi.org/10.1109/TBME.2014.2359372

[9] Catarina Sá, Paulo Veloso Gomes, António Marques and António Correia. 2020. The use of portable EEG devices in development of immersive virtual reality environments for converting emotional states into specific commands. In Multidisciplinary Digital Publishing Institute Proceedings, Vol. 54, No. 1. https://doi.org/10.3390/proceedings2020054043

[10] K. Ezzameli and H. Mahersia. 2023. Emotion recognition from unimodal to multimodal analysis: A review. Information Fusion. https://doi.org/10.1016/j.inffus.2023.101847

[11] Nicu Sebe, Ira Cohen, Theo Gevers, and Thomas S. Huang. 2004. Multimodal approaches for emotion recognition: A survey. Proceedings of SPIE - The International Society for Optical Engineering. https://doi.org/10.1117/12.600746

[12] Guillermo Bernal, Tao Yang, Abhinandan Jain and Pattie Maes. 2018. PhysioHMD: A conformable, modular toolkit for collecting physiological data from head-mounted displays. ACM International Symposium on Wearable Computers (ISWC '18). 160–167. https://doi.org/10.1145/3267242.3267268

[13] Ifigeneia Mavridou, Ellen Seiss, Theodoros Kostoulas, Mahyar Hamedi, Emili Balaguer-Ballester and Charles Nduka. 2019. Introducing the EmteqVR interface for affect detection in virtual reality. International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW '19). 83–84. https://doi.org/10.1109/ACIIW.2019.8925297

[14] Ming-Zher Poh, Daniel J. McDuff and Rosalind W. Picard. 2010. Advancements in noncontact, multiparameter physiological measurements using a webcam. IEEE Transactions on Biomedical Engineering, Vol. 58, No. 1, 7–11. https://doi.org/10.1109/TBME.2010.2086456

[15] Wenjin Wang, Albertus C. den Brinker, Sander Stuijk and Gerard de Haan. 2017. Algorithmic principles of remote PPG. IEEE Transactions on Biomedical Engineering, Vol. 64, No. 7, 1479–1491. https://doi.org/10.1109/TBME.2016.2609282

[16] Xiaobai Li, Jie Chen, Guoying Zhao and Matti Pietikäinen. 2014. Remote-heart-rate measurement from face videos under realistic situations. IEEE Conference on Computer Vision and Pattern Recognition. 4264–4271. https://doi.org/10.1109/CVPR.2014.543

[17] Kwan Long Wong et al. 2022. Optimising rPPG Signal Extraction by Exploiting Facial Surface Orientation. IEEE Conference on Computer Vision and Pattern Recognition Workshop. https://doi.org/10.1109/CVPRW56347.2022.00235

[18] Benjamin Kossack, Eric Wisotzky, Anna Hilsmann and Peter Eisert. 2021. Automatic Region-Based Heart Rate Measurement Using Remote Photoplethysmography. IEEE Conference on Computer Vision and Pattern Recognition Workshop. https://doi.org/10.1109/ICCVW54120.2021.00309

[19] Ru Jing Lee, Saaveethya Sivakumar and King Hann Lim. 2024. Review on remote heart rate measurements using photoplethysmography. Multimedia Tools and Applications, Vol. 83, 44699–44728. https://doi.org/10.1007/s11042-023-16794-9

[20] Shutao Chen et al. 2023. Deep learning-based image enhancement for robust remote photoplethysmography in various illumination scenarios. IEEE Conference on Computer Vision and Pattern Recognition Workshop. https://doi.org/10.1109/CVPRW59228.2023.00647

[21] Quest 2 first step. 2025. Meta official tutorial. [Online]. Available: https://www.meta.com/ko-kr/experiences/first-steps-for-quest-2/3675568169182204/?srsltid=AfmBOopCebqIsBcEtLpqPu1ODQ_g9X291Lvn WxtAV17oX3tGMI1MLmhJ (Accessed on: 2025-05-19)

[22] David Pollreisz and Nima TaheriNejad. 2022. Detection and Removal of Motion Artifacts in PPG Signals. Mobile Networks and Applications, Vol. 27, 728–738. https://doi.org/10.1007/s11036-019-01323-6

[23] CARVE snowboarding. 2025. Meta store page. [Online]. Available: https://www.meta.com/ko-kr/experiences/carve-snowboarding/3317083428365473/ (Accessed on: 2025-05-19)

[24] Affected: The Manor. 2025. Meta store page. [Online]. Available: https://www.meta.com/ko-kr/experiences/affected-the-manor-complete-edition/2967784573246968/ (Accessed on: 2025-05-01)

[25] Fujii. 2025. Meta store page. [Online]. Available: https://www.meta.com/ko-kr/experiences/fujii-a-magical-gardening-adventure/2033429253367669/ (Accessed on: 2025-05-19)

[26] Blair Witch: Oculus Quest Edition. 2025. AltlabVR. [Online]. Available: https://www.altlabvr.com/blair-witch-oculus-quest-edition/ (Accessed on: 2025-05-19)

[27] Wang, X. et al. 2025. Transformer-Driven Affective State Recognition from Wearable Physiological Data in Everyday Contexts. Sensors, Vol. 25, No. 3, 761. https://doi.org/10.3390/s25030761