forum acusticum 2026
8-12 September · Graz, Austria

MOTION ARTIFACT MITIGATION IN AUDIOPLETHYSMOGRAPHY: SIGNAL PROCESSING AND ML-BASED APPROACHES

Dževad Ćoralić[1], Stephan Pranter[1], Fabian Bachl[2]
[1] *USound GmbH, Kratkystraße 2, Graz, Austria*
[2] *USound GmbH, Gutheil-Schoder-Gasse 8-12, Vienna, Austria*

Abstract

Audioplethysmography (APG) enables cardiovascular monitoring through ear-worn devices by detecting cyclical changes to the acoustic properties of the ear canal. However, body-motion-induced artifacts significantly degrade signal quality in real-world applications. This paper investigates the impact of body motion on the accuracy of APG-derived heart-rate estimates and compares a digital signal reconstruction method and three machine-learning-based estimators against an unmitigated baseline. The candidate-ranking model achieved the best overall performance, with a mean absolute heart rate error of 5.36 bpm and a mean absolute percentage error (MAPE) of 6.35%. Its MAPE remained below 5% in all conditions except cycling, where it reached 15.72%. These results show that machine-learning-based candidate selection provides more robust motion-artifact rejection than the signal reconstruction method, although severe body motion remains the primary obstacle to reliable cardiovascular monitoring using APG in everyday scenarios.

Keywords: *audioplethysmography, motion artifacts, signal processing, machine learning, wearable sensors*

1. Introduction

Wearable cardiovascular monitoring continues to be a major research focus in digital health, with applications including heart-rate tracking or blood pressure estimation [1]. With the integration of capable monitoring systems into consumer-facing products, such as smartwatches, development is not limited to medical applications. Currently, many consumer wearables rely on Photoplethysmography (PPG), an optical technique that measures cardiac cycle dependent changes in blood volume within tissue, allowing for the estimation of physiological parameters such as heart rate, pulse waveform, and blood oxygen saturation.While most early systems were wrist-worn, there is increasing interest in ear-worn devices ("earables") [2] , which can unobtrusively monitor physiological signals due to their stable placement and proximity to vascularized tissue in and around the ear canal [3].

1.1. What is Audioplethysmography?

Audioplethysmography (APG) is a novel technique for heart rate measurements based on acoustic sensing of cardiovascular pulse from small cardiac-cycle-dependent changes in the acoustic transfer function of the ear canal [4]. An earphone loudspeaker emits a single ultrasonic probing tone, and an inward-facing microphone records its reflection. Variations in ear-canal geometry and acoustic impedance modulate the carrier; coherent demodulation yields cyclically varying magnitude (R) and phase (Φ) signals. Bilateral sensing therefore provides four APG streams (left and right magnitude and phase) from which heart rate can be estimated.

Unlike Photoplethysmography, APG requires no LEDs or photodiodes. It also differs from in-ear infrasonic hemodynography, which measures low-frequency pressure variations and generally depends on a good seal between earphone and ear canal [5], [6]. APG instead reuses standard earphone's acoustic components, making it attractive for physiological sensing in consumer earables.

1.2. Processing Pipeline

The ultrasonic probing tones are reproduced in the ear through the APG earphones, and the resulting echoes are captured by the built-in microphones. An off-the-shelf USB sound card operating at 96 kHz is used for signal playback and acquisition.

The captured signals are mixed down and decomposed into magnitude (R) and phase (Φ) components,

***Corresponding author**: dzevad.coralic@usound.com.*

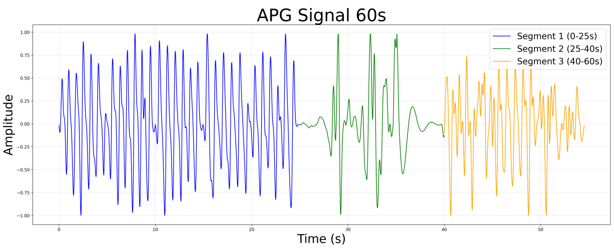


Figure 1: Normalized, concatenated 5-s APG segments under different motion conditions. Blue: clean signal with minimal motion; green: severe artifacts from jaw clenching and rapid head movement, preventing heart-rate extraction; yellow: moderate motion with partial heartbeat periodicity but unreliable heart-rate estimation.

yielding four pulse-wave signals, two per ear. These signals are segmented into 5 s blocks and processed by either the signal-reconstruction or ML-based motion-rejection method to estimate heart rate.
Generally the system operates in two modes: calibration and measurement. During calibration, candidate carrier frequencies from 20 to 39 kHz in 1 kHz steps are evaluated and prioritized according to previous successful selections. Once a suitable probing frequency has been identified for each ear, the system switches to continuous measurement.

2. Challenges in Real-World APG Sensing

Audioplethysmography measurements obtained under controlled laboratory conditions tend to be optimistic: background noise is low, subject movement is restricted, and earphone placement remains stable.

2.1. Motion Artifacts in APG Signals

The desired cardiac modulation is small compared to disturbances caused by speech, chewing, jaw clenching, head motion, exercise, earphone displacement, and cable vibration. As illustrated in Fig. 1, artifacts can change pulse amplitude, create false peaks, remove true peaks, or dominate an entire 5-s analysis block. The problem cannot be solved reliably by simple band-pass filtering because body-motion components often overlap the physiological heart-rate band or appear close to harmonic multiples of the true heart rate. This motivates the two mitigation strategies investigated in this work: reconstructing a plausible periodic waveform from partially usable APG data and learning to reject motion-related or harmonically ambiguous heart-rate hypotheses.

3. Motion Artifact Mitigation Strategies

As an unmitigated baseline, heart rate was estimated directly from each 5–s APG window using the dominant spectral peak in the 0.5–3.5 Hz band, without signal reconstruction or ML-based candidate selection, reporting the channel with the highest spectral contrast.

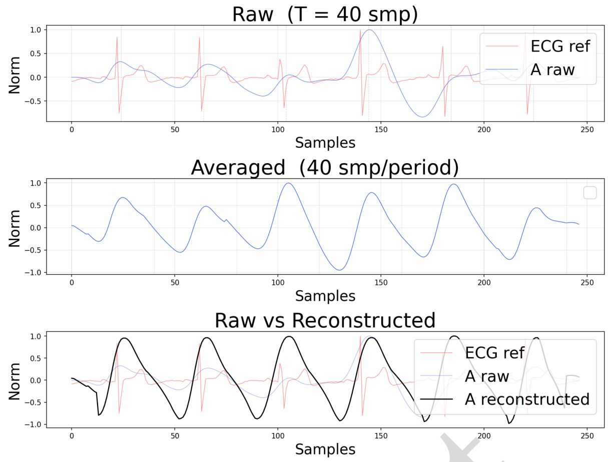


Figure 2: Illustration of the motion rejection and reconstruction process. Top: raw motion-affected signal with the estimated cardiac period. Middle: period-guided averaged signal obtained using interleaving-based smoothing. Bottom: reconstructed signal after pulse correction, shown together with the ECG reference and the original raw recording for comparison.

3.1. Signal Reconstruction Method (SRM)

The signal reconstruction method (SRM) identifies usable segments, estimates the cardiac period, denoises samples at corresponding positions across cycles, and reconstructs corrupted pulses using a template derived from high-quality data. Figure 2 summarizes the process.

3.1.1. Signal usability and period estimation

Each candidate stream, i.e. amplitude and phase, is first evaluated using quantitative quality criteria: a Peak-to-Average Ratio (PAR) (computed to quantify the dominance of periodic components), the number of detected peaks, and the spread of peak-to-peak intervals.
For streams classified as usable, cardiac period is estimated using complementary time- and frequency-domain evidence. First, the signal autocorrelation function is computed using an FFT-based implementation to efficiently capture long-range periodic structure. Autocorrelation is computed over a lag range corresponding to physiologically plausible heart rates (approximately 36-210 bpm). Prominent autocorrelation peaks are refined by quadratic interpolation.
In parallel, a zero-padded FFT identifies the dominant component in the 0.5–3.5 Hz cardiac band. If the estimates agree, the autocorrelation period is preferred; otherwise, their ratio is inspected to correct likely half- or double-period errors. The period is updated whenever subsequent blocks remain valid.

3.1.2. Period-guided averaging / denoising

Following period estimation, the signal is segmented into cycles corresponding to individual cardiac periods. Ideally, each segment captures one complete repetition of the cardiac waveform. For valid signals, the estimated period is stored and later used as a refer-

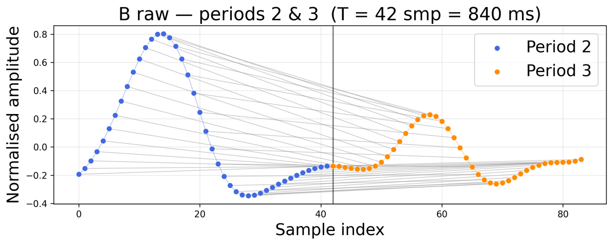


Figure 3: Illustration of the interleaving principle. Two consecutive cardiac periods are shown, with lines indicating samples that occupy the same relative temporal position within each cycle after transposition.

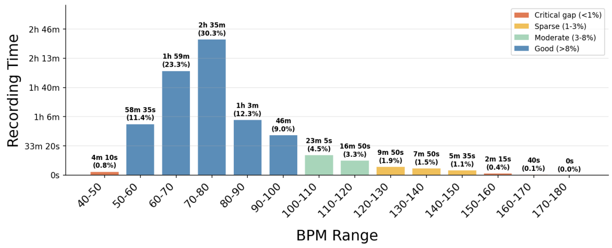


Figure 4: Coverage Analysis - Training data bpm bins by recording time (8h 32m total)

ence to guide the processing of structurally similar but motion-affected signals.

To suppress noise while preserving periodic structure, an interleaving-based filtering technique is applied. The signal is segmented and reshaped into a two-dimensional matrix. After transposition, samples occupying the same relative position in each cycle become adjacent, as shown in Fig. 3. A causal trailing moving-average filter is applied in this interleaved representation, after which the data are de-interleaved. The transformation suppresses non-periodic fluctuations while retaining waveform structure that repeats at the estimated cardiac period.

3.1.3. Pulse template and reconstruction

Peaks and troughs divide accepted cycles into ascending and descending phases. Cycles with invalid ordering or degenerate segments are rejected. Streams with PAR above 2.8 and consistent periodicity contribute to global ascending and descending pulse templates. Extracted pulses are resampled to common lengths and averaged. A motion-affected pulse is then corrected by a convex blend of its local estimate p_{local} and the corresponding global template p_{global},

$$\hat{p} = \alpha p_{\text{local}} + (1-\alpha) p_{\text{global}}, \quad 0 \le \alpha \le 1, \tag{1}$$

with the template contribution increasing as local quality decreases. Beat-to-beat intervals are restricted to 300–2000 ms, and the final channel estimate weights interval- and autocorrelation-based values according to interval stability.

3.2. ML-Based Motion Rejection

To investigate machine learning-based heart rate estimation, a dedicated training dataset was collected from 16 participants under resting and exercise conditions. For every recording, the four demodulated channels were stored together with a fifth channel obtained through blind source separation (BSS) [7]. A Polar H10 [8] ECG chest strap was used as the ground truth device. As Fig. 4 shows, most data lie between 60 and 90 bpm, with comparatively little coverage below 50 and above 100 bpm.

We present three machine-learning approaches to heart-rate estimation, ranging from classical feature-based regression to a fully learned end-to-end architecture. The first directly regresses heart rate from hand-engineered candidate features. The second uses the same features in a learning-to-rank framework to select among physiologically plausible candidates. The third eliminates hand-engineered features, learning frequency-domain representations and temporal dynamics directly from raw signals using convolutional and temporal convolutional networks. Model development used the leave-one-subject-out (LOSO) cross-validation method to measure subject-independent generalization before final comparison on the separate 20-participant held-out test set.

3.3. Candidate Generation and Feature Represantation

The direct-regression and ranking approaches share a candidate-generation front end. In each 5-s window, candidates are extracted independently from all five channels by: (1) spectral-peak detection on the Welch power spectral density; (2) autocorrelation-peak detection; and (3) time-domain peak and inter-beat-interval detection. Every detected value is expanded by 0.5×, 1.5×, and 2× factors. This makes harmonic alternatives explicit rather than requiring the model to discover them indirectly. This step is motivated by the observation, confirmed across all three modelling approaches described in this section, that the dominant source of estimation error is a confusion between the true heart rate and harmonically related aliases, mainly its half-rate sub-harmonic in cases of elevated heart rate or motion.

Each candidate is described by 42 features in six groups: i) candidate value, detector, channel, multiplier, and confidence; ii) channel-level periodicity and quality; iii) agreement with candidates from other channels and detectors; iv) membership in resting or exercise heart-rate ranges; v) waveform morphology such as peak width, symmetry, and amplitude variability; and vi) spectral evidence that the candidate is a harmonic or subharmonic alias.

3.3.1. Direct regression

For each window, the variable-length candidate set is pooled into a fixed 211-dimensional descriptor. Per-feature mean values, maxima, and the three highest-confidence values are combined with the log-transformed candidate count. A Random-Forest regressor predicts the heart rate directly from this window-level representation. The method is a useful baseline because it uses the same candidate evidence as the ranking model but performs neither explicit candidate selection nor temporal modeling. Performance differences therefore isolate the benefit of ranking and decoding rather than differences in the front-end features.

3.3.2. Candidate ranking and temporal decoding

A candidate-pool oracle analysis on the development set, in which the candidate closest to the reference heart rate was selected for each window, yielded an MAE of approximately 1 bpm. This indicates that a near-correct heart-rate candidate is usually present in the pool and that the main remaining challenge is reliable candidate selection. The ranking formulation treats every 5-s window as a group of physiologically plausible hypotheses. An XGBoost pairwise ranking model scores the 42-dimensional candidate vectors using graded labels that favor candidates closest to the ECG reference. A session-level Viterbi decoder then chooses the sequence that balances candidate score with temporal consistency. Large rate jumps are penalized, and additional penalties discourage octave-like transitions associated with half-rate or double-rate switching. This formulation matches the oracle result: APG errors are often selection errors among available candidates, not failures to generate the correct candidate.

3.4. Deep Architecture: Convolutional and Temporal Convolutional Network

The third approach departs from the candidate-generation-and-feature-extraction pipeline shared by the previous two models entirely, learning a heart-rate representation directly from the raw demodulated signal and thereby testing whether a sufficiently expressive learned model can achieve competitive performance without any DSP-derived candidate list or engineered feature vector.

Stage 1: Emission - A compact one-dimensional convolutional network with approximately 113,000 parameters. It processes log-frequency-binned magnitude spectra from the four APG channels and the BSS channel and outputs a posterior distribution over a heart-rate grid with 1-bpm resolution.

Stage 2: Temporal Decoding - A dilated temporal convolutional network (TCN) operating on the sequence of Stage-1 posteriors. The TCN predicts a residual correction to the frozen Stage-1 output; the residual is initialized to zero, so training starts from an identity mapping. A harmonic-penalty loss discourages probability mass near $0.5\times$, $1.5\times$, and $2\times$ the reference rate, while heart-rate-bucket-balanced session sampling reduces bias toward the dominant resting range.

At inference, optional confidence gating replaces isolated low-confidence predictions with interpolation between the nearest sufficiently confident windows. Sessions with no confident anchors are left unchanged. This improves isolated failures but is less useful during sustained high-motion intervals, where interpolation may have no reliable endpoint and can suppress a genuine heart-rate change.

Figure 5: APG measurement setup.

4. Evaluation and Results

4.1. Protocol and Demonstrator

The final evaluation set was recorded from 20 participants under four two-minute conditions: (1) relaxed silence, (2) relaxed listening to music, (3) office work with intensive typing and background music, and (4) pedaling a stationary bicycle. These conditions progressively introduce acoustic activity, upper-body movement, and whole-body motion. Recordings were segmented into 5-s windows, time-aligned with the reference, and processed as continuous sessions. A Polar Verity Sense [9] PPG armband served as the reference modality, selected for its low measurement error [10].

The implementation of APG requires a wide-band ultrasound source while maintaining satisfactory acoustic performance. For this purpose, we use the Greip 1.1 UAM-C1011 two-way speaker module [11] in a custom-built earphone [12]. The design establishes independent acoustic pathways for the woofer and tweeter, allowing the tweeter's high-frequency performance to remain largely unaffected by the geometry required to accommodate the larger woofer.

A wired configuration was selected to simplify the implementation compared with an industry-standard true-wireless architecture.

The experimental setup, shown in Fig. 5, comprises the APG earphones, a dedicated amplifier, a MOTU M4 sound card [13] operating at 96 kHz, and a laptop running custom Python software. A Polar Verity Sense PPG armband [9] is used as the reference sensor.

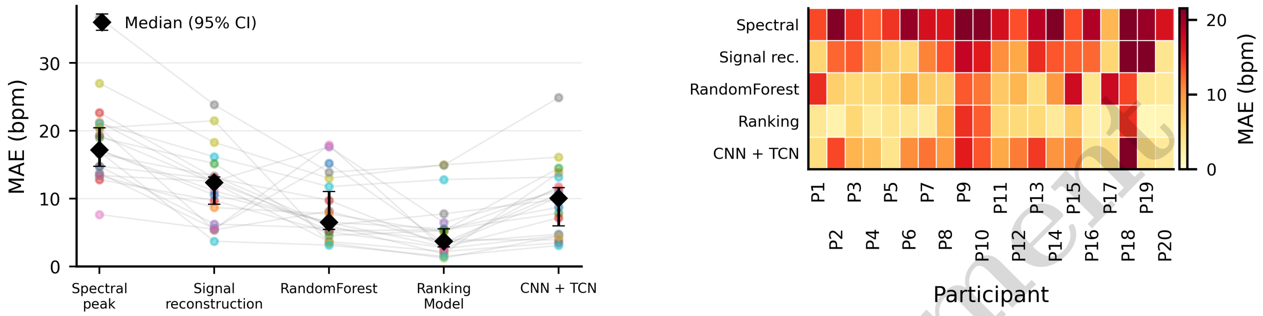
4.2. Overall and Condition-Specific Results

Table 1 reports participant-averaged mean absolute error (MAE) and mean absolute percentage error (MAPE) across all 20 participants for each experimental condition.

The spectral baseline produced the highest overall error (18.31 bpm; 21.91%), confirming that direct spectral-peak selection is unreliable without artifact

Table 1: Participant-averaged MAE (bpm) and MAPE (%) on the held-out test set by experimental condition.

	Silence		Music		Office-work		Biking		Overall	
Method	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
Spectral peak	7.23	10.66	7.85	11.17	16.75	22.87	42.08	43.53	18.31	21.91
Signal reconstruction	4.30	6.69	4.61	6.99	8.44	11.48	30.93	30.71	11.89	13.85
RandomForest	5.88	10.42	5.85	9.90	8.26	13.07	**14.66**	**15.47**	8.55	12.12
Ranking Model	**1.91**	**2.90**	**1.90**	**2.93**	**2.87**	**4.40**	15.40	15.72	**5.36**	**6.35**
CNN + TCN	3.44	5.27	3.75	5.92	5.69	7.90	27.21	27.34	9.78	11.38



(a) (b)

Figure 6: Participant-level MAE: (a) paired results by method; black diamonds and error bars indicate the median and 95% CI; (b) participant–method heatmap.

mitigation. The Ranking Model achieved the lowest overall error (5.36 bpm; 6.35%), followed by RandomForest (8.55 bpm; 12.12%), CNN + TCN (9.78 bpm; 11.38%), and signal reconstruction (11.89 bpm; 13.85%). At the participant level (Fig. 6), the Ranking Model was the most accurate for 13 participants, with 17 participants having an MAE below 10 bpm. Participant P18 was a difficult outlier for all methods (13.9–24.9 bpm MAE), indicating poor recording quality rather than a method-specific failure.

In clinical and scientific evaluations of heart-rate monitoring systems, a mean absolute percentage error (MAPE) below 10% is generally considered acceptable [14]. The Ranking Model and CNN+TCN remained below this threshold in the *Silence*, *Music*, and *Office-work* conditions, with the Ranking Model outperforming CNN+TCN in all four conditions. By contrast, Signal Reconstruction Method and RandomForest exceeded the threshold during *Office-work*. All methods exceeded 10% MAPE during *Biking*, highlighting the impact of severe motion artifacts. For SRM, severe motion can invalidate the period estimate and local pulse waveform required for reconstruction. For CNN+TCN, the limited amount of sustained high-rate, high-motion training data likely reduces the benefit of its learned spectral representation. The Ranking Model remained best overall because of its large advantage in the other three conditions.

4.3. Dependence on Heart-Rate Range

Table 2 groups MAE by reference heart-rate range. All methods are most reliable in the well-represented 50–80 bpm region. The RandomForest error at 40–50 bpm is high (20.57 bpm), consistent with the lack of development data in that range. Errors rise sharply above 100 bpm, a range disproportionately represented by the biking condition, where intense activity, more severe motion artifacts, and sparse training coverage coincide. The 110–120 and 120+ bins contain only 31 windows each; these values reveal an important weakness but should not be treated as precise population estimates.

Table 2: Participant-averaged MAE (bpm) by 10-bpm reference heart-rate bin; n denotes the number of windows.

Range (bpm)	n	Spectral peak	SRM	Random Forest	Ranking Model	CNN + TCN
40–50	115	5.21	8.84	20.57	**5.07**	8.08
50–60	174	8.32	3.19	9.55	**2.84**	3.22
60–70	505	9.89	**3.77**	8.41	4.85	4.47
70–80	321	17.84	7.38	7.65	**7.31**	9.93
80–90	268	29.11	19.26	**7.23**	7.82	17.03
90–100	221	34.45	25.29	**8.79**	9.63	20.04
100–110	82	45.31	41.09	**14.97**	16.25	32.67
110–120	31	56.89	56.34	**22.77**	35.76	44.65
120+	31	75.91	65.46	**49.48**	64.48	69.86

4.4. Participant Variability and Feature Interpretation

Figure 6(a) shows substantial inter-participant variability. Most Ranking Model results cluster at low MAE, while a small number of difficult recordings raise the mean value. The heatmap in Fig. 6(b) shows that failure patterns differ across models: a participant with weak waveform morphology may still benefit from cross-channel consensus, whereas globally corrupted recordings are difficult for every method. Likely contributors include earphone fit, unsuitable probing frequency, heart-rate ranges absent from the training data, and cable-induced artifacts.

5. Conclusion and Key Findings

All mitigation methods improved upon the unmitigated baseline, with candidate ranking and temporal decoding providing the best overall robustness, while Random Forest remained competitive during cycling. The comparison indicates that reliable APG estimation depends on combining cross-channel agreement, explicit harmonic alternatives, physiological context, and temporal consistency. The Signal Reconstruction Method remains useful under low and moderate motion, but methods that explicitly resolve ambiguity between plausible heart-rate candidates are better suited to real-world use.

The presented results highlight the potential of combining APG sensing with ML-based motion artifact mitigation for deployment in consumer audio devices. In particular, executing signal acquisition and preprocessing on a true wireless stereo (TWS) device while performing ML inference on a host device or cloud infrastructure represents a promising deployment strategy. Nevertheless, several challenges remain. Heart rate estimation performance is still influenced by severe motion artifacts and variations in earphone fit, both of which affect signal quality, and even the best-performing method's advantage narrows considerably during vigorous exercise.

5.1. Future Work and Improvements

Future work should broaden the cohort, balance coverage below 50 and above 100 bpm, and include longer recordings under realistic body-motion to better characterize the effects of fit and individual ear-canal transfer functions. Furthermore, the proposed methods should be evaluated under a broader set of motion conditions to better characterize their robustness in real-world usage scenarios.

A further step is transition from the wired demonstrator to an integrated TWS platform. Eliminating cable motion and vibration should reduce mechanically induced artifacts and provide a more representative assessment of real-world performance, while improving signal quality under motion.

Once waveform robustness improves, APG should be evaluated for additional biomarkers such as heart-rate variability, respiratory rate, and blood-pressure-related pulse features.

6. Acknowledgments

The authors would like to thank all study participants and the USound GmbH team for their contributions to data collection, technology development, and manuscript review.

References

[1] H. Xie, L. Yang, B. Jiang, Z. Huang, and Y. Lin, "State-of-the-art wearable sensors for cardiovascular health: A review," *npj Cardiovascular Health*, vol. 2, no. 1, 2025.

[2] K. B. Kim and H. J. Baek, "Photoplethysmography in wearable devices: A comprehensive review of technological advances, current challenges, and future directions," *Electronics*, vol. 12, no. 13, p. 2923, 2023.

[3] L. Clarke, C. Liau, V. W. Wei, and C. Liao, "Healthcare monitoring with earables: A paradigm shift towards digital health," *Biomedical Analysis*, vol. 2, no. 4, pp. 207–222, 2025.

[4] X. Fan, D. Pearl, R. Howard, L. Shangguan, and T. Thormundsson, "APG: Audioplethysmography for cardiac monitoring in hearables," in *Proc. of the 29th Annual International Conference on Mobile Computing and Networking*, 2023.

[5] F. R. Gilliam et al., "In-ear infrasonic hemodynography with a digital health device for cardiovascular monitoring using the human audiome," *npj Digital Medicine*, vol. 5, no. 1, 2022.

[6] R. Ciesielski et al., "Non-invasive blood pressure monitoring with in-ear infrasonic hemodynography for preventative cardiovascular care," *npj Cardiovascular Health*, vol. 2, no. 1, 2025.

[7] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[8] Polar Electro, *Polar H10 Heart Rate Sensor*, Polar Electro.

[9] Polar Electro, *Polar Verity Sense*, Polar Electro.

[10] Y. Liu, A. R. Barker, M. Li, Y. Gu, and C. A. Williams, "Validation of Polar Verity Sense for heart rate monitoring during school-based high-intensity interval training in adolescents," *Journal of Sports Sciences*, vol. 43, no. 11, pp. 1076–1084, 2025.

[11] USound GmbH, *Greip 1.1 UAM-C1011*, `https://usound.com/product/hybrid-audio-module-tws-greip/`, USound GmbH.

[12] USound GmbH, *Greip 1.0 UAM-C1001 acoustic design*, Available upon request, USound GmbH, 2024.

[13] MOTU, *MOTU M4*, MOTU.

[14] B. W. Nelson and N. B. Allen, "Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: Intraindividual validation study," *JMIR mHealth and uHealth*, vol. 7, no. 3, e10828, 2019.