SeatVitals: Geophone-Based Contactless Vital Signs Estimation for Smart Office Health Sensing

Yingjian Song[1,*], Jiayu Chen[1,*], Yida Zhang[1], Zixuan Zeng[1], Zaid Farooq Pitafi[1], Bradley G. Phillips[1], Xiang Zhang[2], Qin Lu[1], Wenzhan Song[1]

[1]University of Georgia, [2]University of North Carolina at Charlotte

Abstract

Real-time monitoring of vital signs in offices enables health promotion, stress management, and workplace optimization. This paper presents SeatVitals, a geophone-based system for unobtrusive and privacy-preserving monitoring of heartbeats and respiration, designed for seamless deployment under office chairs without user involvement. In workplace settings, however, the acquired signals are often heavily contaminated by diverse environmental noise, and labeling heartbeat time series remains inherently challenging. To address these issues, SeatVitals adopts a two-stage learning strategy. To overcome these challenges, SeatVitals first applies self-supervised contrastive learning on large-scale, high-quality sleep data with lower noise levels to learn stable feature representations. The model is then fine-tuned on limited, noisier seated data by aligning these representations with target labels. By leveraging abundant clean unlabeled data together with limited noisy labeled data, SeatVitals achieves both robust generalization and high accuracy in workplace vital sign estimation. Evaluation across 32 subjects yields a mean absolute error (MAE) of 2.44 bpm for heart rate (HR) and 2.20 bpm for respiration rate (RR). The system also achieves competitive performance with only 5% of labeled data for HR and 15% for RR, demonstrating strong label efficiency. These results highlight SeatVitals as a reliable and practical solution for healthcare and wellness monitoring in future smart offices.

1. Introduction

Real-time monitoring of vital signs in workplace settings offers significant potential for promoting health, supporting stress management, and enabling intelligent environmental adaptation in modern smart offices[1, 2]. Existing approaches fall into two categories: wearable devices and contactless methods, but both face limitations when applied in workplace environments.

Wearables provide high-fidelity physiological signals [3], but direct contact can interfere with typing and precision tasks and may be uncomfortable for users with sensory sensitivities. Contactless cameras [4] and mmWave systems [5] improve comfort but raise privacy, installation, and perceived-surveillance concerns that may limit workplace acceptance.

To address these limitations, we present SeatVitals, a seismic sensing system that captures subtle heartbeat vibrations with a geophone placed beneath the chair, enabling contactless monitoring of vital signs in seated scenarios. The design is inherently privacy-preserving and can be seamlessly integrated into workplace environments without imposing burden or discomfort on users, as illustrated in Fig. 1.

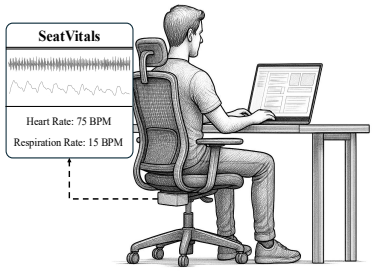


Figure 1: SeatVitals in Working Space

However, the seismic signals collected by the geophone beneath the chair are inherently noisy. Such noise stems not only from the distance and unknown obstructions between the heart and the sensor but also from diverse environmental vibrations in workplace settings, including footsteps and air conditioning. SeatVitals employs a two-step learning framework that transfers knowledge from large-scale, high-quality sleep data to noisy seated data, mitigating workplace noise and enabling robust estimation of HR and RR. In the first stage, self-supervised contrastive learning on a large unlabeled sleep dataset, where signals are cleaner and more abundant, yields stable feature representations. In the second stage, supervised fine-tuning on seated data aligns these representations with target labels. We summarize our contributions as follows:

1. A geophone-based, contactless, and privacy-preserving system for vital sign monitoring in seated scenarios is de-

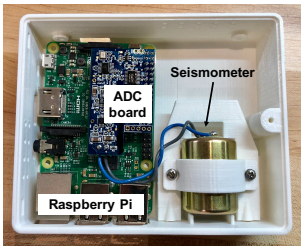
*Equal Contribution.

veloped, enabling real-time and accurate estimation of HR and RR.

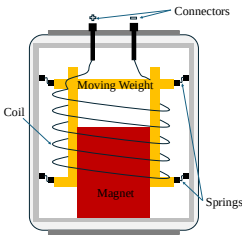
2. A two-step learning framework is proposed, which requires only a small amount of labeled data while overcoming the highly noisy signals in workplace settings.

3. Subject-independent evaluation on 32 seated participants is conducted, where SeatVitals achieves MAE of 2.44 bpm for HR and 2.20 bpm for RR. With only 5% of labeled data for HR and 15% for RR, the system still attains MAEs of 3.25 bpm and 2.22 bpm, respectively.

2. Background and Platform Description



(a) Inside of the SeatVitals



(b) Illustration of the Geophone

As shown in Fig. 2a, SeatVitals consists of a Raspberry Pi, a 100-Hz analog-to-digital converter, and a vertical geophone. The geophone uses a magnet-and-coil structure to convert motion into electrical signals, enabling the capture of subtle cardiac microvibrations (Fig. 2b).

Most existing geophone-based health monitoring systems focused on on-bed scenarios [6–21], where signals generally exhibit higher quality and favorable signal-to-noise ratios. Geophones have also been investigated for activity recognition in office environments [22]. However, seismic sensor-based contactless vital sign estimation in seated scenarios remains largely underexplored.

3. Methodology

3.1. Signals Pre-processing

We refer to geophone-recorded signals as cardiac seismic signals. Because they are measured by seat-mounted sensors far from the heart and are affected by the seat structure, they exhibit high noise and complex, time-varying waveforms. As shown in the first row of Fig. 3, these signals lack clear morphological landmarks and are dominated by harmonics, while the fundamental frequency is often masked by noise. We therefore preprocess each raw signal X^i to extract heartbeat patterns X_h^i and respiration patterns X_r^i.

A 2–12 Hz band-pass filter suppresses high-frequency interference and low-frequency artifacts, yielding clearer heartbeat waveforms (second row of Fig. 3). Because respiration is not directly observable in geophone velocity signals, we first integrate them into displacement, which better captures chest motion, and then apply a low-pass filter to extract respiration patterns (third and fourth rows).

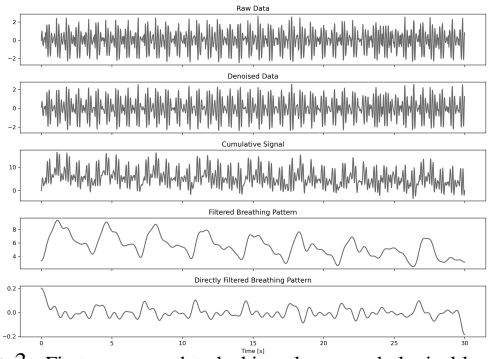


Figure 3: First row: raw data lacking clear morphological landmarks, with continuously varying waveforms. Second row: pre-processed signals for heartbeat patterns. Third and fourth rows: pre-processed signals for respiration patterns. Last row: result of directly applying a lowpass filter to the raw data.

Direct low-pass filtering of the raw velocity signals produces meaningless outputs (last row).

3.2. Self-Supervised Contrastive Learning

As shown in Fig. 4, the encoder and projector are first pre-trained on large-scale unlabeled in-bed data using contrastive learning. During fine-tuning, seated signals are encoded while the regressor is optimized with labeled regression loss and the projector with representation loss for HR and RR estimation.

3.2.1. Data Augmentations

To promote feature invariance and enable robust contrastive learning in the pre-training stage, we apply augmentations that preserve the inherent periodicity of heartbeat and respiration patterns. Specifically, we use: (a) Flipping, which mirrors the input signal horizontally; (b) Scaling, which adjusts the signal amplitude within the range 0.7–1.3. After these transformations, Gaussian noise with zero mean is added.

The extracted heartbeat and respiration signals (X_h^i, X_r^i) are augmented to generate paired views (X_h^{1i}, X_h^{2i}) and (X_r^{1i}, X_r^{2i}). Since HR and RR models are trained independently, both signals are denoted as x_i for simplicity, with paired views expressed as (x_i^1, x_i^2).

3.2.2. Contrastive Representation Learning

To learn robust representations without labels, supervision is provided by constructing positive and negative pairs. Positive pairs are formed from two augmented views of the same input, while negative pairs are formed from different samples.

As illustrated in Fig. 4, two augmented views of samples x_i and x_j are processed by a shared encoder, and the resulting representations are further mapped by a projector into the contrastive embedding space. A contrastive loss is then employed to encourage higher similarity be-

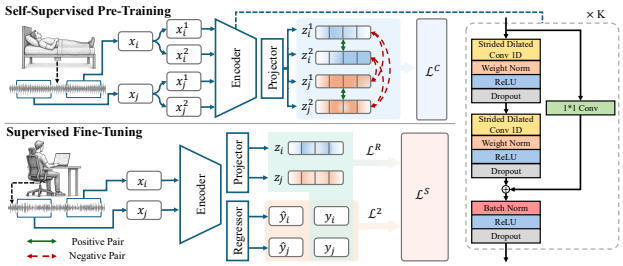


Figure 4: Overview of the proposed two-step framework.

tween embeddings derived from the same input (z_i^1, z_i^2), while reducing similarity with embeddings from other inputs (e.g., z_i^1, z_j^2). Formally, the instance-level loss is defined as $\ell_i^c = -\log \frac{\exp(\text{sim}(z_i^1, z_i^2)/\tau)}{\sum_{j=1}^{2N} \mathbf{1}_{i \neq j} \exp(\text{sim}(z_i, z_j)/\tau)}$ where $\text{sim}(u, v) = \frac{u^T v}{|u| \cdot |v|}$ denotes cosine similarity, and τ is the temperature parameter. The overall contrastive loss $\mathcal{L}^C$ is then averaged across all $2N$ views: $\mathcal{L}^C = \frac{1}{2N} \sum_{i=1}^{2N} \ell_i^c$.

The shared encoder structure, shown in the right panel of Fig. 4, is composed of stacked residual blocks. Each block contains two strided dilated 1D convolution layers with weight normalization, ReLU activation, and dropout, along with a parallel 1×1 convolution to form a residual connection. The outputs are merged via a skip connection, followed by batch normalization, ReLU, and dropout. This structure enables efficient temporal feature extraction while maintaining stability.

3.3. Supervised Representation Learning

Pre-training learns invariant representations, while supervised fine-tuning aligns them with labels to emphasize task-relevant features [23].

To achieve this, the representation loss is introduced to encourage samples with similar labels to lie closer in the latent space. The instance-level loss is defined as:$\ell_{i,j}^r = \| d(z_i, z_j) - |y_i - y_j| \|_2^2$, where y_i and y_j are the ground-truth labels of x_i and x_j, z_i and z_j are their corresponding representations, and $d(\cdot)$ denotes the Euclidean distance. This loss preserves instance-specific features while maintaining consistency with the label space. The overall representation loss $\mathcal{L}^R$ is $\mathcal{L}^R = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \ell_{i,j}^r$

For estimation tasks, a regressor is introduced. The regression loss is defined as $\mathcal{L}^2 = \sum_{i=1}^{N} \|\hat{y}_i - y_i\|_2^2$, where $\hat{y}_i$ denotes the predicted vital sign for sample i, and y_i is the corresponding ground-truth label. This prediction error is incorporated into the fine-tuning objective, yielding the overall supervised loss: $\mathcal{L}^S = \mathcal{L}^2 + \lambda \cdot \mathcal{L}^R$, where λ is an adjustable hyperparameter. During inference, only the encoder and regressor are used.

4. Evaluation

4.1. Evaluation Details

The model was pre-trained on approximately 400 hours of in-bed data from 60 subjects and fine-tuned on 32 seated subjects with approximately 5-minute recordings. FDA-approved CareTaker4 measurements provided HR and RR ground truth. HR and RR were estimated from 10-s and 20-s segments, respectively, using subject-independent train-test splits. Performance was assessed by mean absolute error (MAE), standard deviation (STD), and mean absolute percentage error (MAPE).

4.2. Overall Performance

SeatVitals is compared against the following baselines: (1) HB-phone [8, 9], the first geophone-based system that can accurately monitor the HR, RR in realistic settings; (2) Seismic-based System [7] developed for sleep monitoring; and (3) Fully Supervised Learning, where both the encoder and the regressor are trained end-to-end using all labeled seated data. As shown in Table 1, the proposed two-stage framework consistently outperforms all baselines in MAE, STD, and MAPE for both HR and RR estimation.

Table 1: Performance comparison with baselines. Bold indicates the best performance for each vital sign.

Method	Vital	MAE ↓	MAPE ↓	STD ↓
[8, 9]	HR	7.12	0.103	17.06
	RR	4.48	0.342	3.36
[7]	HR	3.88	0.050	8.82
	RR	3.37	0.243	2.83
SUPERVISED	HR	3.33	0.046	4.71
	RR	2.28	0.190	1.69
PROPOSED	HR	**2.44**	**0.033**	**4.30**
	RR	**2.20**	**0.183**	**1.57**

4.3. Limited Labeled Data Evaluation

We evaluated the proposed model using limited labeled data (5% for HR and 15% for RR) and the full dataset, against two baselines: supervised training on seated data and pre-training on sleep data followed by seated-data fine-tuning. As shown in Fig. 5, our model consistently outperforms both baselines, particularly when labels are scarce. Using only 5% of HR labels or 15% of RR labels, it achieves performance comparable to fully supervised training on all seated data, demonstrating strong label efficiency.

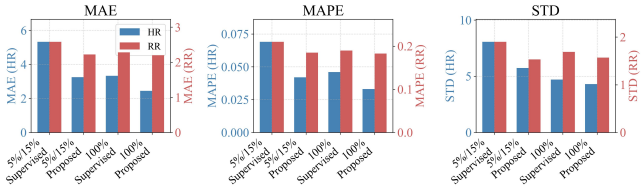


Figure 5: Compare with Fully Supervised Learning

References

[1] Zhang X, Zheng P, Peng T, He Q, Lee CK, Tang R. Promoting employee health in smart office: A survey. Advanced Engineering Informatics 2022;51:101518.

[2] Röcker C. Services and applications for smart office environments-a survey of state-of-the-art usage scenarios. In Proceedings of the International Conference on Computer and Information Technology (ICCIT 2010), Cape Town, South Africa. 2010; 1173–1189.

[3] CarePulse Technology Co., Ltd. Carepulse patch online platform, 2023. URL `https://index.carepulse.cn/home/index.html`. Accessed: September 4, 2025.

[4] Bian H, Guo B, Liu S, Ding Y, Gao S, Yu Z. Ubihr: Resource-efficient long-range heart rate sensing on ubiquitous devices. Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies 2024;8(4):1–26.

[5] Liu J, Wang J, Gao Q, Li X, Pan M, Fang Y. Diversity-enhanced robust device-free vital signs monitoring using mmwave signals. IEEE Transactions on Mobile Computing 2024;23(12):12922–12937.

[6] Park J, Cho H, Balan RK, Ko J. Heartquake: Accurate low-cost non-invasive ecg monitoring using bed-mounted geophones. Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies 2020;4(3):1–28.

[7] Song Y, Li B, Luo D, Xie Z, Phillips BG, Ke Y, Song W. Engagement-free and contactless bed occupancy and vital signs monitoring. IEEE Internet of Things Journal 2023;.

[8] Jia Z, Bonde A, Li S, Xu C, Wang J, Zhang Y, Howard RE, Zhang P. Monitoring a person's heart rate and respiratory rate on a shared bed using geophones. In Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems. 2017; 1–14.

[9] Jia Z, Alaziz M, Chi X, Howard RE, Zhang Y, Zhang P, Trappe W, Sivasubramaniam A, An N. Hb-phone: a bed-mounted geophone-based heartbeat monitoring system. In 2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN). IEEE, 2016; 1–12.

[10] Song Y, Xiang H, Zeng Z, Chen J, Zhang Y, Pitafi ZF, Yang H, Lu Q, Zhang X, Phillips BG, et al. Multi-granularity supervised contrastive learning with online adaptation for contactless in-bed posture classification. Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies 2025;9(2):1–32.

[11] Song Y, Pitafi ZF, Dou F, Sun J, Zhang X, Phillips BG, Song W. Self-supervised representation learning and temporal-spectral feature fusion for bed occupancy detection. Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies 2024;8(3):1–25.

[12] Song Y, Pitafi ZF, Zeng Z, Chen J, Zhang Y, Phillips BG, Brainard BM, Song W. Poster: A contactless health monitoring system for humans and animals. In Proceedings of the 22nd ACM Conference on Embedded Networked Sensor Systems. 2024; 869–870.

[13] Song Y, Li B, Luo D, Glasgow GSB, Phillips BG, Ke Y, Song W. Real-time continuous blood pressure estimation with contact-free bedseismogram. In ICC 2024-IEEE International Conference on Communications. IEEE, 2024; 214–219.

[14] Zhang Y, Chen J, Song Y, Zeng Z, Zhang X, Lu Q, Phillips BG, Xie Z, Song W. Blood pressure estimation from vibration signals via coarse-to-fine contrastive learning, feature selection and synthesis. In Proceedings of the ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies. 2025; 134–138.

[15] Chen J, Song Y, Zhang Y, Zeng Z, Zhang X, Pitafi Z, Xie Z, Das DK, Dong N, Lu J, et al. Selfdenoiser: Self-supervised seismic signal denoiser for continuous and contactless cardiac monitoring. Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies 2025;9(4):1–31.

[16] Song Y. Contactless Sleep Monitoring Via Seismic Sensing. Ph.D. thesis, University of Georgia, 2025.

[17] Li J, Zhang Y, Zeng Z, Chen J, Zhang X, Lu J, Song W, Dou F. Peak-r1: Instruction-tuned large language models for robust j-peak detection in cardiomechanical signals. In NeurIPS 2025 Workshop on Learning from Time Series for Health; .

[18] Li J, Zhang Y, Zeng Z, Chen J, Song Y, Xiao Y, Dong N, Lu J, Kwon Y, Zhang X, et al. Peak-detector: Explainable peak detection via instruction-tuned large language models in physiological signal. Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies 2026; 10(2):1–44.

[19] Song Y, Chen J, Zeng Z, Zhang Y, Pitafi Z, Das DK, Phillips BG, Kwon Y, Healy WJ, Zhang X, et al. Contactless sleep apnea detection with bodyseismography. Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies 2026;10(2):1–31.

[20] Song Y, Chen J, Zeng Z, Zhang Y, Pitafi ZF, Phillips BG, Zhang X, Dou F, Song W. Attention feature fusion with cluster contrastive learning for snoring and breath-holding detection using seismic sensing. In 2026 IEEE International Conference on Pervasive Computing and Communications (PerCom). IEEE, 2026; 1–10.

[21] Pitafi ZF, Yang H, Chen J, Song Y, Ye J, Tse Z, Song K, Song W. A testbed for development and validation of contactless vital signs monitoring systems 2026;.

[22] Bonde A, Pan S, Mirshekari M, Ruiz C, Noh HY, Zhang P. Oac: Overlapping office activity classification through iot-sensed structural vibration. In 2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI). IEEE, 2020; 216–222.

[23] Taghiyarrenani Z, Nowaczyk S, Pashami S, Bouguelia MR. Multi-domain adaptation for regression under conditional distribution shift. Expert Systems with Applications 2023; 224:119907.

Address for correspondence:

Wenzhan Song
University of Georgia, Athens, GA 30602, USA
wsong@uga.edu