EarCalo: Earable-Based Energy Expenditure Estimation While Running

Yang Liu
yl25r@fsu.edu
Florida State University
Tallahassee, FL, USA

Qiang Yang
qy258@cam.ac.uk
University of Cambridge
Cambridge, UK

Jake Stuchbury-Wass
js2372@cam.ac.uk
University of Cambridge
Cambridge, UK

Dong Ma
dm878@cam.ac.uk
University of Cambridge
Cambridge, UK

Cecilia Mascolo
cm542@cam.ac.uk
University of Cambridge
Cambridge, UK

Abstract

Accurately estimating energy expenditure (EE) is crucial for understanding exercise efficiency, managing fitness goals, and monitoring health conditions. Existing wearable systems either rely primarily on heart rate and motion sensors, resulting in unsatisfactory accuracy, or require bulky setups such as thermal cameras to improve performance by integrating more physiological information, which limits real-world applicability. We present EarCalo, an earable-based system that leverages in-ear audio sensing to estimate EE during running. The system extracts airflow-induced acoustic variations within the ear canal and employs a deep neural network to translate these subtle in-ear sound dynamics into EE estimates. The key insight is that in-ear acoustic signals can capture multiple physiological factors such as breathing and cardiovascular activity, while also reflecting motion-related cues like running intensity. These rich acoustic cues are closely related to EE and provide a unified sensing modality for estimation. We evaluated EarCalo on 21 participants running at varying speeds using a mixed-user setting. EarCalo achieved a mean absolute error (MAE) of 0.67 kcal/min, a mean absolute percentage error (MAPE) of 11.98%, and a Pearson correlation of 0.945, which achieves accuracy close to established physiological standards. This work represents an early step toward practical and personalized earable-based EE estimation in everyday settings.

CCS Concepts

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

Keywords

In-ear audio, wearable sensing, energy consumption estimation

ACM Reference Format:
Yang Liu, Qiang Yang, Jake Stuchbury-Wass, Dong Ma, and Cecilia Mascolo. 2026. EarCalo: Earable-Based Energy Expenditure Estimation While Running. In *The 27th International Workshop on Mobile Computing Systems and Applications (HotMobile '26), February 25–26, 2026, Atlanta, GA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3789514.3792046

ACM ISBN 979-8-4007-2471-8/2026/02
https://doi.org/10.1145/3789514.3792046

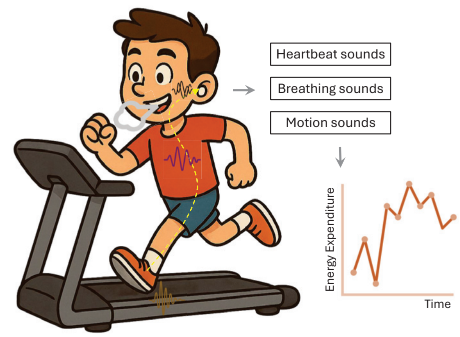

Figure 1: EarCalo uses the in-ear microphones of earphones to capture physiological and motion-related sounds for energy expenditure monitoring.

1 Introduction

Energy expenditure (EE) refers to the amount of energy consumed by the body to maintain essential physiological functions and perform physical activities [9, 27]. EE generally includes resting metabolism, the energy required for food digestion, and the energy spent during voluntary movement or exercise [9]. Among these components, exercise-related EE is the most dynamic and directly reflects how the body responds to physical activity [14]. Accurate estimation of EE during exercise is crucial for understanding workout efficiency, managing fitness goals, and assessing health conditions such as obesity, diabetes, and cardiovascular disease [14, 32].

Existing EE estimation solutions span laboratory-grade methods to consumer-grade wearables. Direct calorimetry, which measures heat output, provides gold-standard accuracy but requires sealed chambers and is unsuitable for real-world exercise [20]. Indirect calorimetry, which estimates EE from oxygen consumption and carbon dioxide production, offers practical reference accuracy but depends on bulky respiratory masks and gas analyzers [1, 20]. Consumer-grade wearables such as accelerometers and photoplethysmography (PPG) sensors on smartwatches are compact and accessible but often inaccurate. Studies show that such devices can deviate by more than 40% from true EE values due to their limited physiological sensing scope and reliance on heuristic models [2, 4, 10, 13]. Recent advances [2] integrate thermal imaging with respiration sensing to improve EE estimation accuracy during cycling and running. However, these systems still rely on cameras that are bulky, sensitive to lighting and occlusion, and raise privacy concerns.

Motivated by the need for accurate yet practical energy expenditure monitoring, we explore earable sensing as a new opportunity for this task. The ear is a compact and stable location that is already instrumented in everyday life through earphones, providing a natural platform for unobtrusive physiological sensing [5, 7, 8, 16–18, 21, 30, 33]. We present EarCalo, the first earable-based system that leverages in-ear audio sensing to estimate EE during running. The system captures airflow-induced acoustic variations inside the ear canal and uses a deep neural network to map these subtle sound dynamics to energy expenditure. The key intuition is that in-ear acoustics inherently encode multiple physiological processes that are directly linked to EE. Prior studies have demonstrated that in-ear audio can capture respiratory airflow, cardiovascular pulsation, and motion-induced vibrations [6, 22, 23, 31, 34, 35]. These signals respectively reveal oxygen uptake and metabolic rate, reflect blood flow and oxygen transport, and indicate the mechanical effort exerted by the body [20]. By integrating these complementary factors through a single, compact modality, EarCalo provides a unified view of physical exertion for accurate EE estimation.

To evaluate this idea, we collected data from 21 participants running at different speeds using in-ear microphones synchronized with indirect calorimetry (VO_2 Master Analyzer mask [1]) as the ground truth. We trained a general model across all participants to capture diverse acoustic and physiological patterns and evaluated it on unseen data from these participants. Our analysis demonstrates that EarCalo achieves a MAE of 0.67 kcal/min, a MAPE of 11.98%, and a Pearson correlation of 0.945 with calorimetry-based EE estimates. A MAPE below 10% is typically considered satisfactory [19, 25, 28], and our results approach this standard. Overall, this work represents an early step toward enabling practical and personalized earable-based EE estimation in everyday settings.

In summary, this work makes the following contributions:

- We present EarCalo, the first system to investigate in-ear acoustic sensing for estimating energy expenditure during running. We recognize that in-ear acoustics provide a unified channel capturing multiple physiological processes closely related to EE.
- We develop a deep learning model that translates subtle ear-canal acoustic dynamics into accurate energy expenditure estimates validated against indirect calorimetry.
- We evaluate EarCalo through a 21-participant study using a mixed-user setting, demonstrating that our system can achieve accuracy close to established standards while maintaining comfort, mobility, and everyday usability.

2 Background

2.1 Energy Expenditure

Energy expenditure (EE) quantifies the amount of energy the body uses to sustain life and perform physical activity. It reflects an individual's overall metabolic state and is commonly represented as the **total energy expenditure (TEE)** [9]. TEE comprises several physiological components:

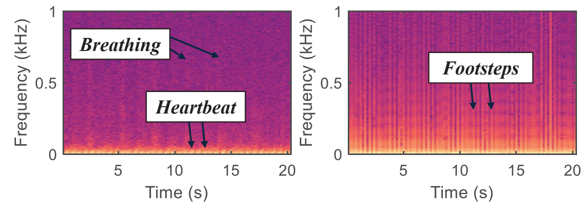

Figure 2: Spectrograms of in-ear audio.

- **Resting Energy Expenditure (REE):** the baseline energy required to maintain vital physiological functions such as circulation, respiration, and cellular metabolism.
- **Diet-Induced Energy Expenditure (DEE):** the energy consumed to digest, absorb, and metabolize nutrients.
- **Activity-Related Energy Expenditure (AEE):** the additional energy spent on voluntary physical activity, varying widely depending on activity type, intensity, and duration.

During exercise, total energy expenditure increases substantially as the body's metabolic demand rises to supply oxygen and nutrients to active muscles and to maintain thermoregulation. Reliable **TEE estimation under exercise** provides a direct measure of physical workload and metabolic efficiency [14, 32]. It enables users to understand how their bodies respond to different exercise intensities, manage calorie balance, and design personalized training programs. Beyond personal fitness, it supports clinical assessment of metabolic health, rehabilitation progress, and cardiovascular function [14, 32]. At a broader scale, large-scale and continuous EE tracking can inform studies on population health, lifestyle behavior, and the prevention of metabolic diseases.

2.2 Energy Expenditure from the Ear

The ear presents an ideal sensing location for monitoring EE due to the following reasons.

Physiological accessibility. The ear is anatomically close to key physiological pathways, including the carotid artery and upper airway, enabling access to rich biosignals such as heart rate, heartbeat output, and respiration [6, 15, 22, 23, 31]. In-ear microphones benefit from the occlusion effect, allowing the detection of subtle acoustic variations caused by airflow and vascular pulsations [6, 31]. These signals are directly linked to oxygen intake, cardiac output, and metabolic rate.

Motion sensing. In-ear microphones can also capture motion-related cues during exercise as body vibrations propagate through the ear canal via bone conduction [12, 22–24]. These acoustic variations reflect physical activity intensity.

We validated these observations as shown in Figure 2. The left figure shows in-ear spectrograms when the user is stationary with moderate breathing. Distinct frequency patterns corresponding to breathing cycles and heartbeat pulses confirm that in-ear microphones capture both respiratory and cardiovascular activities. When the user begins running (right), rhythmic footstep patterns dominate, indicating that the in-ear signal also reflects motion intensity.

Relevance to energy expenditure. Energy expenditure is driven by multiple physiological and behavioral processes, including respiration rate, cardiovascular effort, and body motion [20]. The ear uniquely provides access to all three, allowing a single

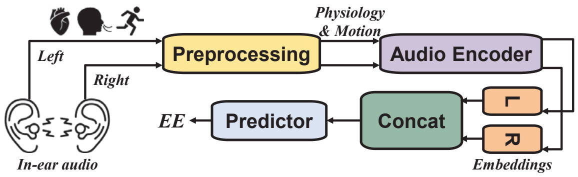

Figure 3: EarCalo architecture.

sensing location to capture multimodal cues that reflect the body's overall exertion level.

By leveraging in-ear audio sensing, EarCalo unifies respiration, cardiovascular, and motion information within a compact and practical wearable form factor, namely earbuds that are already widely adopted and provide a natural, unobtrusive sensing platform for EE estimation.

3 System Design

3.1 Overview

EarCalo estimates energy expenditure (EE) from in-ear acoustic signals through the pipeline shown in Figure 3. The system takes synchronized left- and right-channel in-ear audio as input and produces an EE estimate as output. Each audio segment is first preprocessed to emphasize relevant physiological and motion information, then encoded into high-level acoustic embeddings through an audio encoder. These embeddings capture respiratory and cardiovascular modulations as well as motion-induced vibrations associated with running intensity. A convolutional neural network (CNN) and multilayer perceptron (MLP) jointly serve as the predictor, transforming the embeddings into an EE value representing the caloric expenditure rate (kcal/min).

3.2 Preprocessing

Raw in-ear audio is segmented into fixed-length windows (12s) without overlap. Each segment is transformed into a compact time–frequency representation as follows:

- **Log-Mel spectrogram extraction:** We compute log-Mel spectrograms with 64 Mel bins spanning 50 Hz–8 kHz using a 1024-point window and a 320-point hop size. This configuration effectively preserves the fine-grained physiological and motion variations encoded in in-ear acoustics: frequencies near 50 Hz correspond to cardiac pulsations [6], the 50–150 Hz band reflects footstep and body-vibration components [22], and higher frequencies capture airflow turbulence and harmonic structures of respiration.
- **Channel pairing:** Spectrograms from the left and right in-ear microphones are temporally aligned and stacked to form a synchronized two-channel input tensor.

This preprocessing preserves fine-grained acoustic variations associated with respiration, cardiovascular pulsations, and motion-induced vibrations, providing rich multimodal cues for downstream EE estimation.

3.3 Audio Encoder

To obtain robust and transferable representations from limited labeled data, EarCalo fine-tunes a Contrastive Language–Audio Pretraining (CLAP) encoder [11] pretrained on large-scale sound datasets. CLAP converts each time–frequency input into a 1024-dimensional embedding, capturing temporal and spectral cues of the in-ear audio. Separate embeddings are extracted from the left and right channels and concatenated into a 2×1024 matrix representing both ear-specific and shared acoustic dynamics.

During fine-tuning, the encoder is adapted to the in-ear acoustic domain and optimized to align its feature representations with the patterns underlying energy expenditure. This process adjusts the pretrained feature space to emphasize acoustic variations that correlate with physical exertion level, enabling EarCalo to generate representations that are discriminative for EE estimation while preserving the generalization benefits of large-scale pretraining.

3.4 Energy Expenditure Predictor

The concatenated embeddings are passed through a lightweight CNN followed by MLP layers to estimate EE:

- **Convolutional fusion:** The first 2D convolution uses a kernel size of $(2, 3)$, spanning both ear channels and three neighboring embedding dimensions. This operation performs early channel fusion while learning local feature correlations that reflect subtle differences in spectral and physiological characteristics encoded by the embeddings.
- **Hierarchical encoding:** A subsequent convolution with kernel size $(1, 3)$ further abstracts local dependencies along the embedding dimension while preserving the fused channel representation. Each convolutional layer is followed by ReLU activation, batch normalization, and max pooling to enhance stability and generalization.
- **Regression head:** The resulting feature map is flattened and passed through a compact multilayer perceptron with one hidden layer and dropout regularization, projecting the learned representation into a single scalar output corresponding to EE (kcal/min).

The model is trained end-to-end using mean squared error (MSE) loss between the predicted EE $\hat{E}_i$ and the ground-truth calorimetry value E_i:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^{N} (E_i - \hat{E}_i)^2, \tag{1}$$

where N is the number of training samples. This objective encourages accurate per-window EE estimation while allowing the convolutional stack to learn stable inter-channel and and feature-space relationships in the embeddings.

4 Evaluation

4.1 User Study

To evaluate EarCalo, we conducted a controlled treadmill study with 21 healthy participants. The in-ear microphones were synchronized with a VO_2 Master Analyzer mask [1], which served as the ground truth (GT) device. All procedures were approved by the institutional ethics committee, and informed consent was obtained from all participants.

Custom Earbuds. As commercial Active Noise Cancellation (ANC) earphones do not provide access to raw in-ear microphone data, we developed a custom hardware prototype for data collection. Figure 4 (top left) shows the device, which consists of a pair of earphones embedded with miniature in-ear microphones (Knowles

Figure 4: Experimental setup. Participants wore custom earphones with in-ear microphones and a VO_2 Master mask for reference EE measurement while running on a treadmill.

SPU1410LR5H-QB) facing the ear canal to capture acoustic signals. The microphones were connected to a Raspberry Pi 4 equipped with a custom PCB and audio codec board, housed in a lightweight chest pouch for portability. Audio was sampled at 44.1 kHz.

Ground Truth (GT) Device. For ground truth, we used the VO_2 Master Analyzer mask [1] (Figure 4, bottom left), a mobile indirect calorimetry device that estimates oxygen uptake and carbon dioxide output, enabling real-time EE calculation (kcal/min). The mask was calibrated before each session using a 3 L reference syringe. Synchronization between the earphones and GT data was performed manually by aligning the timestamps of distinct deep-breathing events recorded before running.

Procedure. Each recording session lasted approximately 10 minutes of treadmill running (Figure 4, right). Participants ran at two self-selected speeds: a comfortable pace (light jog) and a faster pace (moderate to high exertion), each for 5 minutes. This design induced varying EE levels. No specific breathing rhythm or running style was imposed, ensuring that the dataset captured natural, unconstrained physiological behavior.

Participants. We recruited 21 participants, including 9 female and 12 male, aged 23–53 years (*mean*=29.0, *SD*=6.9), with body mass indices (BMI) ranging from 18.0 to 30.8 kg/m^2. Several participants were regular runners, while others had limited exercise habits, providing a diverse range of aerobic fitness levels and metabolic responses. The running speeds ranged from 3 km/h to 12 km/h (*mean*=6.35, *SD*=1.82), covering light to vigorous running conditions. In total, we collected near 210 minutes of synchronized in-ear audio and ground-truth EE data. The calorimetry-derived EE values ranged from 0.67 to 17.6 kcal/min across participants and conditions, reflecting different running intensities. The diversity in individual physiology and exercise levels provides a comprehensive dataset for training and evaluating EarCalo.

4.2 Implementation

We trained and evaluated EarCalo under a mixed-user setting, where audio segments from all participants were pooled together and randomly divided into 80% for training and 20% for testing. This setting allows the model to learn general acoustic–EE patterns across users while still evaluating on unseen segments to assess generalization. Each audio segment was labeled with its corresponding energy expenditure value obtained from the ground truth device.

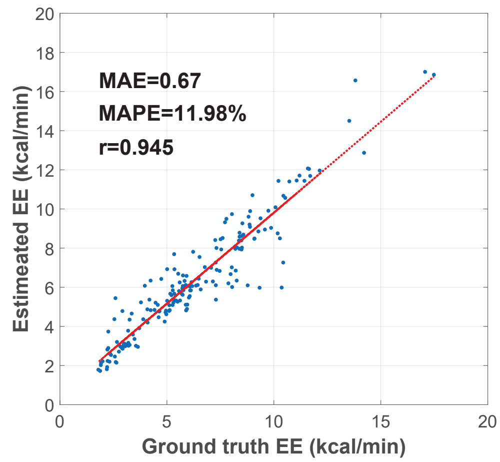

Figure 5: Performance of EarCalo on unseen test segments under mixed-user evaluation setting.

In practice, a user could obtain several minutes of personalized calibration data (e.g., through a brief supervised session in a clinic or gym equipped with a calorimetry reference device) to fine-tune the model for individual physiology. This adaptation would enable more accurate long-term EE estimation tailored to each user.

All models were implemented in PyTorch using the 2023 version of CLAP [11] as the pretrained audio encoder. The network was optimized using Adam with a learning rate of 1×10^{-5}, batch size of 32, and dropout ratio of 0.1.

4.3 Performance

Evaluation Metrics. We evaluated EarCalo using three standard regression metrics that quantify estimation accuracy and consistency with the ground-truth EE values:

- *Mean Absolute Error (MAE):* the average absolute difference between predicted and reference EE values (kcal/min).
- *Mean Absolute Percentage Error (MAPE):* the average absolute error divided by the ground truth (%).
- *Pearson Correlation Coefficient (r):* the linear correlation between predicted and reference EE, where $r = 1$ indicates perfect agreement.

Overall Performance. Figure 5 shows the overall performance of EarCalo. The system achieves a MAE of 0.67 kcal/min and a MAPE of 11.98% when evaluated on unseen test segments under mixed-user evaluation setting. Prior studies [19, 25, 28] consider a MAPE below 10% to be satisfactory for energy expenditure estimation, and our results are close to this level of accuracy. The scatter plot illustrates the relationship between predicted and reference EE values, revealing a strong linear trend with a Pearson correlation coefficient of $r = 0.945$. The estimates closely follow the ground truth across the full EE range, with only a few isolated outliers. Furthermore, the errors are evenly distributed around the identity line, suggesting that the model does not exhibit systematic overestimation or underestimation tendencies.

5 Discussion and Future Work

Model and representation design. EarCalo shows that in-ear acoustics contain rich information reflecting respiration, heartbeat, and motion dynamics, all of which are closely related to energy expenditure. However, the current model processes these components together without explicitly distinguishing their physiological origins. Future work will investigate architectures that can separate and integrate these complementary sources more effectively. For instance, multi-branch or attention-based models could individually encode respiratory, cardiovascular, and motion-related features and then combine them through a learned fusion module. This design would allow the network to emphasize the relative contribution of each physiological factor, leading to more interpretable and accurate estimation.

Activity scope. The current evaluation focuses on treadmill running, which provides controlled conditions with stable environments and consistent motion. Real-world exercise, however, spans a wider range of activities and movement contexts. Extending the study to other exercise types such as walking, cycling, stair climbing, high-intensity interval training, and weight training will help assess how in-ear acoustics reflect energy expenditure across different intensities and exercise forms. This will also test the robustness of EarCalo under outdoor variability (e.g., posture changes, ambient noise, and diverse individual exercise styles). Our future work will extend the design and evaluation to outdoor settings and a broader activity set, and report performance across intensities to characterize failure modes and generality.

Participant Diversity and Generalization. Our study included 21 healthy adults covering a wide range of ages (23 to 53 years), BMI (18.0 to 30.8 kg/m^2), and fitness levels (both regular runners and non-runners). In the current implementation, we trained and evaluated EarCalo using a mixed-user setting, where audio segments from all participants were pooled and randomly divided into training and testing sets. This configuration mainly evaluates the model's overall predictive capability rather than its ability to generalize to unseen individuals.

Physiological responses to exercise and the resulting acoustic signatures can vary considerably across people, particularly among older adults or those with respiratory or cardiovascular conditions. To gain an initial understanding of cross-user generalization, we conducted a preliminary leave-one-subject-out (LOSO) evaluation, which yielded a MAE of 1.59 kcal/min and a MAPE of 25.57%. These results show that while the mixed-user model captures general patterns effectively, its performance decreases when applied to unseen users, reflecting the challenge of inter-individual variability. Future work will involve recruiting a larger and more diverse participant cohort to further study this variability and explore methods that balance generalization and personalization. Incorporating user demographic attributes such as age, sex, or BMI into the model may also help explain physiological differences and improve generalization across users. In addition, domain adaptation and meta-learning techniques may enhance cross-user robustness, while lightweight user-specific calibration using a few minutes of labeled data could further refine individual performance.

Robustness in in-the-wild conditions. In-the-wild operation introduces additional confounders for in-ear audio sensing, including environmental noise (e.g., wind/traffic), earbud fit variation due to sweat and motion, and concurrent audio playback. Although the occluded ear canal provides partial isolation from ambient sound, fit changes and playback leakage can still distort physiological cues. Our future work will explicitly stress-test these factors and improve robustness via (i) noise/playback-aware augmentation and training, and (ii) signal-quality estimation to detect poor-fit/noisy segments.

Systems considerations on COTS earables. Practical deployment on commodity earbuds requires an end-to-end systems design that balances accuracy with latency, energy, memory, and communication overhead. We plan to profile the full pipeline under two representative implementations: on-earbud inference and earbud-to-phone offloading. This includes quantifying the trade-off between streaming raw audio versus transmitting compact features/embeddings, and measuring the corresponding energy/latency costs. We will also move toward a fully wireless prototype to capture realistic communication overhead and evaluate adaptive policies (e.g., activity- and quality-aware sampling/compute) that reduce battery impact while maintaining accuracy.

6 Related Work

Laboratory-grade Methods. Accurate measurement of energy expenditure (EE) traditionally relies on calorimetry-based methods. Direct calorimetry measures the body's heat output within a sealed chamber, providing gold-standard precision but being impractical for everyday use [20]. Indirect calorimetry, which estimates EE based on oxygen consumption and carbon dioxide production, provides reliable reference accuracy but requires bulky equipment such as respiratory masks and gas analyzers [1, 20]. These systems are therefore restricted to laboratory or clinical environments rather than real-world exercise scenarios.

Consumer-grade Wearables. Modern wearables such as Apple Watch and Fitbit, as well as existing research prototypes, typically estimate EE using accelerometers and PPG sensors [2]. While lightweight and accessible, these methods rely on activity-dependent or heuristic models and often exhibit high inter-device variability and limited accuracy [2, 4, 10, 13]. Multisensor approaches, such as combining an Inertial Measurement Unit (IMU) with ECG [26] or deploying multiple IMUs on different body parts [29], can improve accuracy but usually require several electrodes or sensors, reducing comfort and everyday usability. Other studies have explored laser- and camera-based sensing of chest motion [3], but these systems remain sensitive to body movement. Recent work such as JoulesEye [2] leverages wearable thermal imagery to capture respiration for EE estimation during running and cycling, demonstrating that integrating richer physiological cues can improve performance. However, such systems still require users to keep the camera facing the face, which increases effort and raises usability and privacy concerns in daily exercise contexts.

EarCalo explores a new sensing locus and modality using in-ear acoustics captured by earbuds. It naturally fits mobile exercise scenarios where users already wear earphones, and the ear canal provides a compact, privacy-preserving site that simultaneously captures respiration, cardiovascular pulsations, and motion-induced vibrations, enabling unified inference from a single, everyday form factor.

7 Conclusion

This paper presents an early exploration into a practical and reliable wearable-based approach for estimating energy expenditure during exercise. We introduced EarCalo, an earable-based system that estimates energy expenditure during running by leveraging airflow- and vibration-induced acoustic variations inside the ear canal. To our knowledge, this is the first work to demonstrate that earphones can support accurate energy expenditure estimation. The system offers a compact and natural form factor that integrates respiratory, cardiovascular, and motion information for metabolic estimation. Our 21-participant study shows that EarCalo achieves a MAPE of 11.98%, which achieves accuracy close to established physiological standards. These findings highlight the potential of earable devices as a physiologically grounded and unobtrusive platform for practical energy expenditure monitoring. Future work will explore more expressive model architectures, a wider range of exercise scenarios, and adaptive methods for addressing inter-user variability.

Acknowledgements

This research was supported by ERC project 833296 and EPSRC grant EP/Z53447X/1. The research was conducted during the first author (Y.L.)'s appointment at the University of Cambridge.

References

[1] 2025. VO2 Master. https://vo2master.com/.
[2] Rishiraj Adhikary, Maite Sadeh, Nipun Batra, and Mayank Goel. 2024. Jouleseye: Energy expenditure estimation and respiration sensing from thermal imagery while exercising. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 4 (2024), 1–29.
[3] Hirooki Aoki and Hidetoshi Nakamura. 2018. Non-contact respiration measurement during exercise tolerance test by using kinect sensor. *Sports* 6, 1 (2018), 23.
[4] Rob Argent, Megan Hetherington-Rauth, Julie Stang, Jakob Tarp, Francisco B Ortega, Pablo Molina-Garcia, Moritz Schumann, Wilhelm Bloch, Sulin Cheng, Anders Grøntved, et al. 2022. Recommendations for determining the validity of consumer wearables and smartphones for the estimation of energy expenditure: expert statement and checklist of the INTERLIVE network. *Sports Medicine* 52, 8 (2022), 1817–1832.
[5] Jacob Brown, Yang Liu, and Cecilia Mascolo. 2023. Yawning Detection using Earphone Inertial Measurement Units. In *Proceedings of the 2nd Workshop on Smart Wearable Systems and Applications*. 7–13.
[6] Kayla-Jade Butkow, Ting Dang, Andrea Ferlini, Dong Ma, Yang Liu, and Cecilia Mascolo. 2024. An evaluation of heart rate monitoring with in-ear microphones under motion. *Pervasive and Mobile Computing* 100 (2024), 101913.
[7] Justin Chan, Antonio Glenn, Malek Itani, Lisa R Mancl, Emily Gallagher, Randall Bly, Shwetak Patel, and Shyamnath Gollakota. 2023. Wireless earbuds for low-cost hearing screening. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services*. 84–95.
[8] Tao Chen, Yongjie Yang, Xiaoran Fan, Xiuzhen Guo, Jie Xiong, and Longfei Shangguan. 2024. Exploring the feasibility of remote cardiac auscultation using earphones. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 357–372.
[9] James P DeLany and Jennifer C Lovejoy. 1996. Energy expenditure. *Endocrinology and metabolism clinics of North America* 25, 4 (1996), 831–846.
[10] Peter Düking, Laura Giessing, Marie Ottilie Frenkel, Karsten Koehler, Hans-Christer Holmberg, Billy Sperlich, et al. 2020. Wrist-worn wearables for monitoring heart rate and energy expenditure while sitting or performing light-to-vigorous physical activity: validation study. *JMIR mHealth and uHealth* 8, 5 (2020), e16716.
[11] Benjamin Elizalde, Soham Deshmukh, and Huaming Wang. 2024. Natural language supervision for general-purpose audio representations. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 336–340.
[12] Andrea Ferlini, Dong Ma, Robert Harle, and Cecilia Mascolo. 2021. EarGate: gait-based user identification with in-ear microphones. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 337–349.
[13] Daniel Fuller, Emily Colwell, Jonathan Low, Kassia Orychock, Melissa Ann Tobin, Bo Simango, Richard Buote, Desiree Van Heerden, Hui Luan, Kimberley Cullen, et al. 2020. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. *JMIR mHealth and uHealth* 8, 9 (2020), e18694.
[14] Andrew P Hills, Najat Mokhtar, and Nuala M Byrne. 2014. Assessment of physical activity and energy expenditure: an overview of objective measures. *Frontiers in nutrition* 1 (2014), 5.
[15] Changshuo Hu, Thivya Kandappu, Yang Liu, Cecilia Mascolo, and Dong Ma. 2024. Breathpro: Monitoring breathing mode during running with earables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (2024), 1–25.
[16] Changshuo Hu, Thivya Kandappu, Jake Stuchbury-Wass, Yang Liu, Anthony Tang, Cecilia Mascolo, and Dong Ma. 2024. Detecting Foot Strikes during Running with Earbuds. In *Proceedings of the Workshop on Body-Centric Computing Systems*. 35–40.
[17] Changshuo Hu, Qiang Yang, Yang Liu, Tobias Röddiger, Kayla-Jade Butkow, Mathias Ciliberto, Adam Luke Pullin, Jake Stuchbury-Wass, Mahbub Hassan, Cecilia Mascolo, et al. 2025. A Survey of Earable Technology: Trends, Tools, and the Road Ahead. *arXiv preprint arXiv:2506.05720* (2025).
[18] Tamisa Ketmalasiri, Yu Yvonne Wu, Kayla-Jade Butkow, Cecilia Mascolo, and Yang Liu. 2024. IMChew: Chewing Analysis using Earphone Inertial Measurement Units. In *Proceedings of the Workshop on Body-Centric Computing Systems*. 29–34.
[19] Shenglong Le, Xiuqiang Wang, Tao Zhang, Si Man Lei, Sulin Cheng, Wu Yao, and Moritz Schumann. 2022. Validity of three smartwatches in estimating energy expenditure during outdoor walking and running. *Frontiers in Physiology* 13 (2022), 995575.
[20] James A Levine. 2005. Measurement of energy expenditure. *Public health nutrition* 8, 7a (2005), 1123–1132.
[21] Jiao Li, Yang Liu, Tao Sun, Ziheng Zhou, and Jin Zhang. 2025. Earable-based Continuous Blood Pressure Monitoring via a Single-Point Flexible Sensor. In *Companion of the 2025 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 828–833.
[22] Yang Liu, Kayla-Jade Butkow, Jake Stuchbury-Wass, Adam Pullin, Dong Ma, and Cecilia Mascolo. 2025. Respear: Earable-based robust respiratory rate monitoring. In *2025 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 67–77.
[23] Yang Liu, Qiang Yang, Kayla-Jade Butkow, Jake Stuchbury-Wass, Dong Ma, and Cecilia Mascolo. 2025. EarMeter: Continuous Respiration Volume Monitoring with Earables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 4 (2025), 1–29.
[24] Dong Ma, Andrea Ferlini, and Cecilia Mascolo. 2021. Oesense: employing occlusion effect for in-ear human sensing. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*. 175–187.
[25] M Benjamin Nelson, Leonard A Kaminsky, D Clark Dickin, and ALEXANDER H Montoye. 2016. Validity of consumer-based physical activity monitors for specific activity types. *Medicine and science in sports and exercise* 48, 8 (2016), 1619–1628.
[26] Zhiqiang Ni, Tongde Wu, Tao Wang, Fangmin Sun, and Ye Li. 2022. Deep multi-branch two-stage regression network for accurate energy expenditure estimation with ECG and IMU data. *IEEE Transactions on Biomedical Engineering* 69, 10 (2022), 3224–3233.
[27] R Passmore and Jm VGA Durnin. 1955. Human energy expenditure. *Physiological reviews* 35, 4 (1955), 801–840.
[28] Lilian Roos, Wolfgang Taube, Nadja Beeler, and Thomas Wyss. 2017. Validity of sports watches when estimating energy expenditure during running. *BMC Sports Science, Medicine and Rehabilitation* 9, 1 (2017), 22.
[29] Patrick Slade, Mykel J Kochenderfer, Scott L Delp, and Steven H Collins. 2021. Sensing leg movement enhances wearable monitoring of energy expenditure. *Nature Communications* 12, 1 (2021), 4312.
[30] Jake Stuchbury-Wass, Yang Liu, Kayla-Jade Butkow, Josh Carter, Qiang Yang, Mathias Ciliberto, Ezio Preatoni, Dong Ma, and Cecilia Mascolo. 2025. WalkEar: holistic gait monitoring using earables. In *2025 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 99–109.
[31] Jordan Waters, Jake Stuchbury-Wass, Yang Liu, Kayla-Jade Butkow, and Cecilia Mascolo. 2025. Deep-learning based segmentation of in-ear cardiac sounds. In *2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 1–7.
[32] Klaas R Westerterp. 2013. Physical activity and physical activity induced energy expenditure in humans: measurement, determinants, and effects. *Frontiers in physiology* 4 (2013), 90.
[33] Qiang Yang, Yang Liu, Jake Stuchbury-Wass, Kayla-Jade Butkow, Dong Ma, and Cecilia Mascolo. 2024. BrushBuds: Toothbrushing Tracking Using Earphone IMUs. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 655–660.
[34] Qiang Yang, Yang Liu, Jake Stuchbury-Wass, Kayla-Jade Butkow, Emeli Panariti, Dong Ma, and Cecilia Mascolo. 2025. SmarTeeth: Augmenting Manual Toothbrushing with In-ear Microphones. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19.
[35] Qiang Yang, Yang Liu, Jake Stuchbury-Wass, Mathias Ciliberto, Tobias Röddiger, Kayla-Jade Butkow, Adam Luke Pullin, Emeli Panariti, Dong Ma, and Cecilia Mascolo. 2025. HearForce: Force Estimation for Manual Toothbrushing with Earables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 4 (2025), 1–22.