Materials Research Proceedings **27** (2023) 300-307 https://doi.org/10.21741/9781644902455-39

Integration of wearable and ambient sensors towards characterization of physical effort

Aaron Appelle[1,a *], Liming Salvino[2], Yun-An Lin[3], Taylor Pierce[3], Emerson Noble[3], Gabriel Draughon[2], Kenneth J. Loh[3], and Jerome P. Lynch[1]

[1] Duke University, Durham, NC 27708, USA

[2] University of Michigan, Ann Arbor, MI 48109, USA

[3] University of California San Diego, La Jolla, CA 92093, USA

[a]aaron.appelle@duke.edu

Keywords: Physiological Monitoring, Multimodal, Synchronization, Thin Film, Kinesiology Tape, Geophone, Microphone, Computer Vision, Signal Processing

Abstract. Human performance monitoring in complex operational environments calls for sensing solutions that measure human physiology as well as human interactions with their surroundings. Recent advances in multimodal sensing have led to the development of intelligent environments that analyze human activities with high granularity. One of the greatest challenges is to unify multiple discrete sensing systems through synchronization and integration of multimodal data streams. This paper describes an intelligent environment that consolidates wearable skin-strain sensors for physiological monitoring; geophones and microphones to record ambient vibrations and sounds; and video cameras to visually observe human activities. We show proof-of-concept functionality by using the system to differentiate walking effort in human subjects. First, the work shows the alignment of wearable and ambient sensor time-history records. Then, data features are extracted and correlated to walking speed using three sensor modalities. Finally, feature-level analysis is done to associate the data features with the perceived walking exertion for each subject.

Introduction

Wearable devices that measure physiological signals such as heart rate, skin temperature, electrodermal activity (EDA), and electrocardiogram (ECG) are increasing in popularity. Wrist-worn devices offer satisfactory monitoring capabilities when an overall metric of well-being is desired, for example, fitness tracking in sports applications [1], fatigue monitoring of construction workers [2], and performance evaluation of military populations [3]. However, physiological signals provide only limited information about the activities or surroundings of the person wearing the device. When perception of the activities in an environment takes precedent over physiological monitoring, one may instead embed sensors into the environment itself: so-called intelligent spaces or environments [4, 5]. These intelligent environments aim to understand what happens within using visual, depth, and motion sensors. Relatively less work has attempted to combine the paradigms of wearable and ambient sensing to monitor human activity and performance at a fine-grained level. Such unified sensing solutions would benefit team operations where there is a clear interplay between environmental conditions and personal well-being (e.g., military operations, construction, and industrial operations).

Intelligent environments with different sensor types face the challenge of integrating multiple heterogeneous data sources to create unified analyses of human performance. In general, wearable and ambient sensors are not time-synchronized. Multimodal unification is commonly achieved using a feature-level data fusion approach. In contrast to raw data fusion where different sensor

(a) Motion Tape (b) GS-14-L9 Geophone (c) IMP23ABSU Microphone (d) GW5037IP Camera

Figure 1: Overview of sensors used as part of the intelligent environment.

records are combined on a sample-by-sample basis, feature-level fusion requires only rough alignment since data features are extracted from the raw data before being combined into feature vectors [6].

In this paper, an intelligent environment is proposed to consolidate ambient sensing with wearable sensing for the purpose of human performance monitoring. The ambient sensors integrated into the environment are video cameras, geophones, and microphones. The wearable sensors are graphene "motion tape" skin-strain sensors [7] capable of monitoring muscle engagement and respiration [8]. The system is used in a validation case study to characterize perceived walking effort (i.e., pace). Two volunteers are instructed to walk at different speeds (which correlate to subject exertion effort) ranging from slow to fast while being monitored by the multisensory intelligent environment. From the video cameras, computer vision techniques are employed to estimate the subject walking velocity. The geophones, which measure ground vibrations to reveal footstep impacts, are used to compute subject step rates. Motion tape sensors mounted on the chests of subjects are used as respiration monitors where sensor signals contain information on the rate and duration of breaths. The features from the three different sensing modalities are then combined to associate each trial with the perceived walking effort. The results inspire future human subject monitoring experiments aimed at activity detection and quantification of physical exertion through unified wearable and ambient sensing.

System Hardware

Wearable sensors: motion tape is a new type of skin-strain sensor manufactured by spray-coating athletic kinesiology tape with a graphene nanosheet to give it piezoresistive properties [7]. This study uses motion tape placed on a chest band to monitor breathing (Fig. 1a). The manufacturing process permits flexible shapes and sizes for the sensor, which results in a varying resistance range for each sensor. The sensors connect to a custom-designed wearable data acquisition node which keeps time based on its internal oscillator but synchronizes across nodes by connecting them to a base-station computer at the start of the experiment. Motion tape data is acquired at a sampling frequency of 20 Hz.

Ambient sensors: geophones and microphones record ambient ground vibrations and sounds, respectively. Geophones are sensors that convert the velocity of motion into voltage. They can be used to reveal the vibration response of footstep ground impacts. The geophone of choice is a GS-14-L9 from Geospace Technologies (Fig. 1b). Raw geophone signals are passed through a 10-150 Hz band-pass filter for anti-aliasing and a 2000 times amplification to improve the signal strength prior to analog-to-digital conversion. The microphone is an omnidirectional micro-electromechanical systems (MEMS) microphone from STMicroelectronics (Fig. 1c). It is small enough to integrate into printed circuit boards. It features a flat, extended frequency response up to 80 kHz for recording sounds through the ultrasound range. Raw audio signals are passed through a 28 Hz–78.5 kHz band-pass filter and onboard amplification of 360 times. All geophones and microphones are sampled at 100 kHz; this high rate is intended to take advantage of the ultrasound range of the microphones. A National Instruments PXI modular data acquisition system (DAQ) is

used for data collection. The DAQ synchronizes the geophones and microphones using its real-time operating system. Video of the environment and subjects is recorded using a standard 1080p IP camera (Fig. 1d) using a 24 frames per second frame capture rate. Video footage is streamed to a base station computer using Real Time Streaming Protocol (RTSP) and backed up to a Network Video Recorder (NVR).

In summary, four sensor types (motion tape, geophone, microphone, camera) and three different data acquisition platforms are utilized to create an intelligent environment with elements of both physiological monitoring and ambient environmental sensing. Each DAQ records a UTC timestamp to ensure system-wide synchronization. It is assumed that the timestamp labels are accurate down to the second. Post-processing alignment of multimodal sensor feeds from the three DAQ platforms begins by extracting the starting timestamp of each DAQ record and shifting time histories relative to one another to get sub-second resolution. Alignment is manually verified by observing prominent signal features corresponding to impulsive events like subject stomping, clapping, or running which are visible in all sensors time histories.

Experiment

The sensing system is implemented in a validation study at Mcity [9], an outdoor smart city laboratory at the University of Michigan in Ann Arbor. The laboratory's open outdoor space and information technology (IT) infrastructure enable comprehensive human subject tests. In this study, the focus is on single-person walking tests with two different subjects as a case-study to demonstrate the integration of wearable and ambient sensors. Subject 1 is asked to walk with two different walking efforts: "slow" pace in Test 1, which is close to a natural walking pace, and "moderate" pace in Test 2, which is brisk but not jogging. Subject 2 is asked to walk with three different efforts: "slow" pace in Test 1, "moderate" pace in Test 2, and "jogging" pace in Test 3, where moments of the stride have both feet off the ground. The trajectories of the subjects are visualized in Fig. 2. It is noted that future holistic testing could recruit a larger number of subjects and use a documented effort scale such as Borg's Rating of Perceived Exertion (RPE) [10], which has widely been used to rank the difficulty of tasks in other physiological studies.

The goal of the experiment is to determine the walking effort in both subjects by unifying data features from the wearable motion tape sensors with features from the ambient geophone and visual sensors. In this specific work, the microphones are primarily used to assist with data alignment. The alignment of all sensors is illustrated in Fig. 3, taken from Subject 1, Test 2. Several important physiological and activity features appear in the data. The top two data streams show the walker's (x, y) coordinates, computed using computer vision techniques from the video feed (see section *Methodology* for more details). The next four data streams come from the geophones which reveal transient oscillations for each footstep. As the walker approaches then leaves the vicinity of each geophone, the envelope of the vibrations increases then decreases in amplitude correspondingly. The amplitude swells first at Geophone 1 then 2, 3, and 4, indicating that the subject is walking counterclockwise. The data from the motion tape sensor on the subject's chest clearly shows peaks for each inhale occurring at approximately once every two seconds. The motion tapes on the left and right sides of the subject's abdomen also reflect walking motion. The frequency of peaks in each abdominal record is approximately half of the step rate (visible in the geophone data). Each time the left leg steps forward, a peak in the left abdomen sensor occurs, and vice-versa for the right abdomen. The sum of the peaks from the left abdomen and right abdomen (20 peaks each) is equal to the number of footsteps visible in the geophone feeds (40 oscillations).

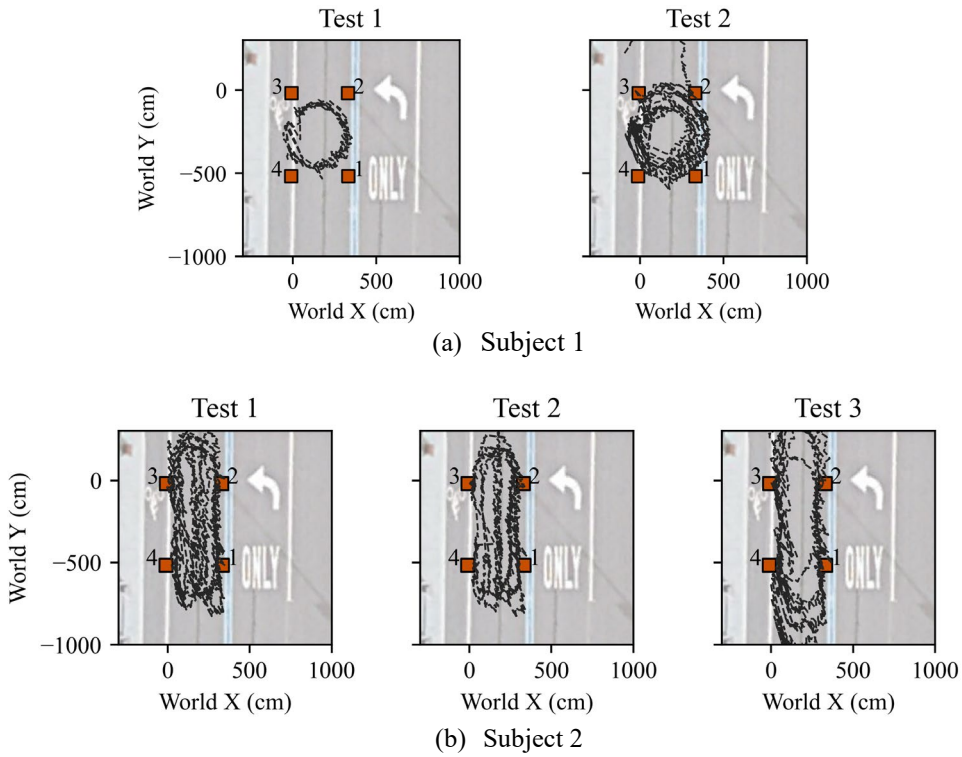


Figure 2: Visualization of routes taken by Subject 1 and Subject 2 at Mcity. Red squares numbered 1 through 4 represent the locations of the sensor boxes containing co-located microphones and geophones. The coordinate system is chosen by the authors with the origin at sensor box number 3. The video camera (not shown) is located at $(x, y) = (-484, 28.2)$ *cm.*

Methodology

Motion tape: using the chest-worn motion tape sensors, two data features are computed: subject respiration rate, measured in breaths per second, and normalized signal power. The respiration rate is computed as the frequency corresponding to the peak of the motion tape data in the frequency domain. The fast Fourier transform (FFT) algorithm [11] is used to compute the discrete Fourier transform (DFT) of the data. Although the breathing rate is generally expected to increase with physical effort, it is also influenced by the rhythm of the physical activity itself. Historical studies have shown the entrainment of breathing rate with the rhythm of motion [12], which is not rigidly tied to the exercise intensity. For this reason, an additional signal power feature is computed which aims to account for the intensity of breaths. The data is pre-processed using a digital low-pass filter with a 2 Hz cutoff frequency to isolate breathing motion. The signal is normalized to have maximum and minimum values of ± 1 and a mean of zero. Then, the signal power P_x is computed as the signal energy per unit time: $P_x = \frac{1}{N}\sum_{n=0}^{N-1} |x(n)|^2$ for a discrete signal $x(n)$ of length N. An illustration of the signal power feature is given in Fig. 4, where the more profound peaks in the jogging signal result in higher signal power than the slow walking signal.

Geophones: the geophones are used to compute a step rate, in steps per second, by counting the number of footstep events per unit of time. The event detection algorithm first implemented by Pan et al. [13] is used with minor modifications. The algorithm intakes windowed vibration time series data and determines whether the current window represents ambient vibration or a footstep event. First, the acquired discrete signal $x(n)$ of length N is windowed every 0.1 seconds in increments of 20 msec. Next, a signal energy feature $\mathrm{E_x} = \sum_{n=0}^{N-1} |x(n)|^2$ is computed. The signal

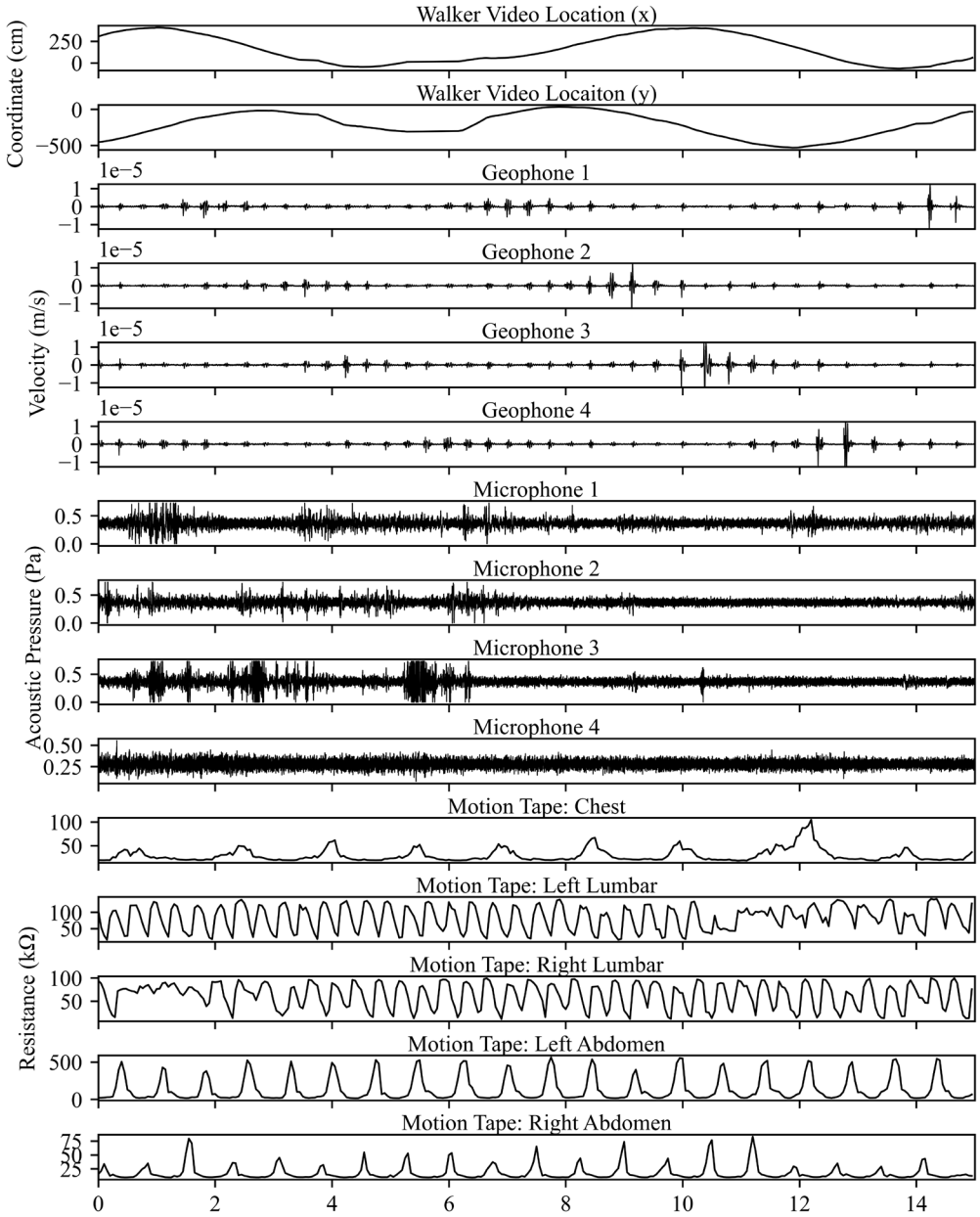


Figure 3: Alignment of multimodal sensor data. Subject 1 location (top two plots), geophones (next four plots), microphones (middle four plots), and motion tape sensors (bottom five plots)

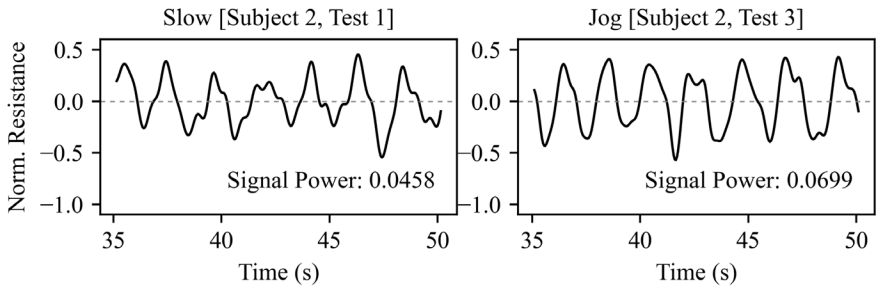


Figure 4: Signal power feature on normalized motion tape chest data, different efforts.

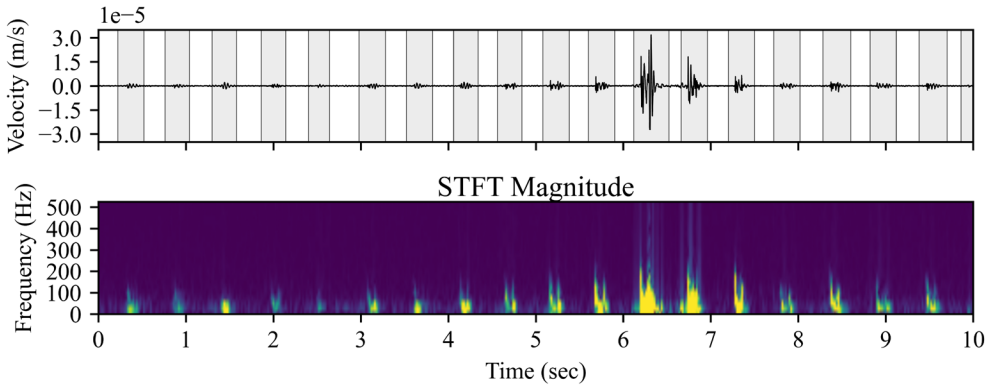
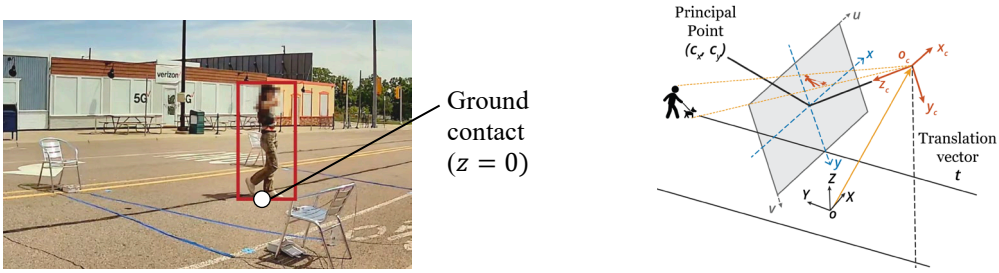


Figure 5: Top: Geophone signal with footstep events in shaded windows, computed using signal energy method. Bottom: time-frequency representation of geophone signal using the short-time Fourier transform (STFT).

energy is compared to a Gaussian noise model $X \sim N(\mu, \sigma^2)$ whose parameters are continuously updated based on non-footstep-event data. If the current window's energy feature exceeds three standard deviations (3σ) above the mean energy of the noise (μ), then the window is labeled a *potential* step event. The *potential* step event is re-labeled as a *confirmed* step event if the next window also exceeds the $\mu + 3\sigma$ threshold, and if at least two out of the four geophones label it as such. Finally, consecutive windows labeled as footstep events are merged. Visualization of the algorithm is shown in Fig. 5 where shaded windows of the plot indicate footstep events. The step rate is the number of steps divided by the duration of the record. Also shown in Fig. 5 is the short-time Fourier transform (STFT), showing that footstep impulses registered by the geophones primarily consist of the frequencies below 200 Hz after signal conditioning.

Video: The video camera is used to estimate an average walking velocity of the person in the intelligent environment (limited to a single person in this study). The YOLO object detector [14] detects the person in each video frame and draws a bounding box surrounding the subject (Fig. 6a). It is assumed that the detected person is standing on the ground, which represents a height coordinate of $z = 0$ in the world coordinate system. Successive coordinate transformations as described by the *pinhole camera model* (Fig. 6b) are used to convert from pixel coordinates of the base of the YOLO bounding box to 3D coordinates in a world system. The inverse projection is made tractable by the $z = 0$ assumption. Since location estimates using this method can jitter frame-by-frame, a 12-point (0.5 seconds) simple moving average (SMA) is applied to the coordinates extracted. Once the coordinates of the monitored subject are computed for each video frame, the instantaneous walking velocity is estimated as the change in Euclidean distance from one frame to the next divided by the time between video frames. The data feature is the average walking velocity over all video frames in each test.



(a) Person detection using YOLO (b) Pinhole camera model for 3D reconstruction [15]

Figure 6: Person localization using computer vision from video feed.

Results

Fig. 7 shows the four data features computed across tests for the two subjects walking with differing levels of exertion. Step rate and walking speed show similar patterns in line with expectations: both subjects exhibit faster step rates and higher average walking speed when told to walk with higher effort. Individual differences are also evident between the two subjects. Subject 1 walking with moderate effort uses a step pace similar to Subject 2's jogging effort (Fig. 7a), however the speed is not as fast (Fig. 7b), indicating shorter stride length. Subject 1 breathes faster walking with moderate speed than slow speed (Fig. 7c). This is also true of Subject 2, but Subject 2 breathes more slowly while jogging than walking moderately. We attribute the reduction in respiration rate to entrainment with running pace. Subject 2's jogging step rate is 2.6 steps/second and respiration rate 0.5 breaths/second; therefore, the subject breathes approximately once every five steps. On the other hand, the signal power data feature for Subject 2 increases according to perceived effort, while that is not true for Subject 1. It appears that increased effort results in faster, shallower breaths by Subject 1 but slower, deeper breaths by Subject 2. The overall impression of the four multimodal data features agrees with subjects' perceived walking effort. Step rate and walking speed show identical trends that can pick out the walking rate despite person-specific breathing patterns. The breathing features give an understanding of the adaptations to breath rate and depth that subjects make with higher physical effort.

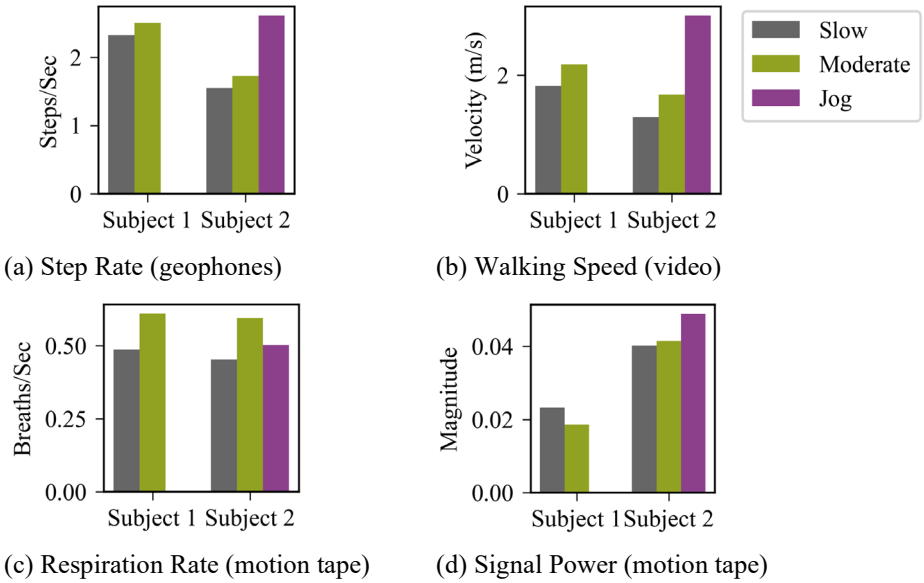


Figure 7: Walking features computed across sensor modalities. Sensor type indicated in parentheses. Perceived walking effort by the subject is color-coded.

Conclusion

This paper advances recent efforts to integrate physiological and environmental monitoring to create unified metrics of human performance. Using video cameras, geophones, and wearable motion tape skin-strain sensors, four data features related to perceived walking exertion are computed. The data features are shown to accurately reflect walking effort in a case study from an outdoor validation experiment with two volunteers. Next steps in quantifying physical effort in an intelligent environment could involve controlled human subject experiments where each volunteer is asked to perform repetitions of an activity at different RPE levels. Additional physiological sensors that measure heart rate and/or ECG would complement the respiration measurement by chest-worn motion tapes. Since the work herein shows proof of concept, the research team is

actively pursuing more advanced human performance monitoring tasks using the same multimodal sensor suite. Other tasks include localization of multiple people and detection of actions and poses.

Acknowledgements
This research was supported by the U.S. Office of Naval Research (ONR) under grants N00014-22-1-2095 and N00014-21-1-2033 (principal investigator: Jerome P. Lynch) and N00014-20-1-2329 and N00014-22-1-2550 (principal investigator: Kenneth J. Loh). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the supporting funding agencies.

References
[1] G. Aroganam, N. Manivannan, and D. Harrison, "Review on Wearable Technology Sensors Used in Consumer Sport Applications," *Sensors*, vol. 19, no. 9, p. 1983, Jan. 2019. https://doi.org/10.3390/s19091983
[2] A. Aryal, A. Ghahramani, and B. Becerik-Gerber, "Monitoring fatigue in construction workers using physiological measurements," *Autom. Constr.*, vol. 82, pp. 154–165, Oct. 2017. https://doi.org/10.1016/j.autcon.2017.03.003
[3] K. Hinde, G. White, and N. Armstrong, "Wearable Devices Suitable for Monitoring Twenty Four Hour Heart Rate Variability in Military Populations," *Sensors*, vol. 21, no. 4, Art. no. 4, Jan. 2021. https://doi.org/10.3390/s21041061
[4] J.-H. Lee and H. Hashimoto, "Intelligent Space — concept and contents," *Adv. Robot.*, vol. 16, no. 3, pp. 265–280, Jan. 2002. https://doi.org/10.1163/156855302760121936
[5] G. Draughon, J. Lynch, and L. Salvino, "Integrated Vision-Body Sensing System for Tracking People in Intelligent Environments," in *European Workshop on Structural Health Monitoring*, Cham, 2023, pp. 885–893. https://doi.org/10.1007/978-3-031-07254-3_89
[6] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proc. IEEE*, vol. 85, no. 1, pp. 6–23, Jan. 1997. https://doi.org/10.1109/5.554205
[7] Y.-A. Lin *et al.*, "Graphene K-Tape Meshes for Densely Distributed Human Motion Monitoring," *Adv. Mater. Technol.*, vol. 6, no. 1, p. 2000861, 2021. https://doi.org/10.1002/admt.202000861
[8] Y.-A. Lin, E. Noble, C.-H. Loh, and K. J. Loh, "Respiration Monitoring using a Motion Tape Chest Band and Portable Wireless Sensing Node," *J. Commer. Biotechnol.*, vol. 27, no. 1, p. 32, Jan. 2022. https://doi.org/10.5912/jcb1026
[9] "Mcity," *Mcity*. https://mcity.umich.edu/ (accessed Nov. 24, 2022).
[10] G. A. V. Borg, "Psychophysical bases of perceived exertion," *Med. Sci. Sports Exerc.*, vol. 14, no. 5, pp. 377–381, 1982. https://doi.org/10.1249/00005768-198205000-00012
[11] J. W. Cooley and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series," *Math. Comput.*, vol. 19, no. 90, pp. 297–301, 1965. https://doi.org/10.1090/S0025-5718-1965-0178586-1
[12] R. R. Bechbache and J. Duffin, "The entrainment of breathing frequency by exercise rhythm.," *J. Physiol.*, vol. 272, no. 3, pp. 553–561, Nov. 1977. https://doi.org/10.1113/jphysiol.1977.sp012059
[13] S. Pan, A. Bonde, J. Jing, L. Zhang, P. Zhang, and H. Y. Noh, "BOES: Building Occupancy Estimation System using sparse ambient vibration monitoring," in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2014*, Apr. 2014, vol. 9061, pp. 406–421. https://doi.org/10.1117/12.2046510
[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91
[15] P. Sun, G. Draughon, R. Hou, and J. P. Lynch, "Automated Human Use Mapping of Social Infrastructure by Deep Learning Methods Applied to Smart City Camera Systems," *J. Comput. Civ. Eng.*, vol. 36, no. 4, p. 04022011, Jul. 2022. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000998