Emotion Recognition Using Footstep-Induced Floor Vibration Signals

YUYAN WU, YIWEN DONG, SUMER VAID,
GABRIELLA M. HARARI and HAE YOUNG NOH

ABSTRACT

Structural vibrations induced by human footsteps contain rich information that can be used for a wide range of applications, including occupant identification, localization, activity recognition, and health and emotional state estimation. Among these, emotion recognition holds great potential for improving smart buildings by enabling mental health monitoring and human-centric services. Existing emotion recognition approaches use cameras, wearables, and mobile devices to capture people's changing gait patterns under various emotional states. However, these approaches come with corresponding drawbacks, such as being limited by visual obstructions and requiring users to carry devices that cause discomfort.

To overcome these drawbacks, we introduce a new emotion recognition approach using footstep-induced structural vibration signals. The main intuition of this approach is that people's gait patterns change under various emotions [1], thus inducing distinct structural vibration patterns as they walk. Compared to other methods, our approach is non-intrusive, insensitive to visual obstructions, and has fewer perceived-privacy concerns. The main research challenge in developing our approach is that emotions have both explicit and implicit effects on gait, making the explicit gait parameters insufficient to describe such a complex relationship. To this end, we develop a set of emotion-sensitive features from the vibration signals, including gait parameters, sequential features, and time-frequency spectrum features to capture both the explicit and implicit effects of emotion on gait. To better integrate multiple types of features, we develop the fully-connected layer, the long-short-term-memory (LSTM) layer, and the convolutional layer to extract information from the features and a multilayer perceptron to estimate emotion. Our approach is evaluated in a real-world walking experiment involving 5 participants with over 100 minutes of footstep-induced floor vibration signals. Our results show that our approach achieves a mean absolute error of 1.33 for valence score estimation and 1.26 for arousal score estimation out of an overall score range of 1 to 9, which has an accuracy of 72% for High / Low valence classification and 82% for High / Low arousal for emotion classification.

Yuyan Wu, PhD Student, Email: wuyuyan@stanford.edu. Structures as Sensors Lab (SaS Lab), Depart- ment of Civil and Environmental Engineering, Stanford University, Stanford, CA, USA.

INTRODUCTION

Emotion recognition refers to the process of identifying human emotions, which has significant and diverse applications, such as monitoring mental health and enabling smart home devices to dynamically adjust their responses and functionalities to better accommodate the user's emotional needs. It can aid individuals in identifying their emotional states, potentially decreasing the risk of depression and anxiety [2]. It can also be used to enhance recommendation systems in smart homes and smart building devices. By being able to infer and interpret human emotions, smart homes and smart building devices can better understand and respond to human behavior [3–5]. Additionally, emotion recognition can help target emotion-based advertisements, resulting in more effective and personalized marketing [6–8].

In recent years, researchers have utilized various aspects of human behaviors to recognize human emotions, such as facial expressions [9, 10], body language [11–14], physiological signals (e.g., EEG, skin conductance) [15], and speech [16]. The sensors used to capture these behaviors typically include cameras, wearable devices, mobile devices (e.g., smartphones), microphones, and RF devices. However, each of these sensing methods has its own limitations. For instance, cameras are susceptible to obstructions and have limited viewing angles, which also raise privacy concerns. Similarly, microphones can raise privacy concerns and require the subject to speak, and RF devices may be blocked or absorbed by objects such as water or metal. Wearable or mobile devices require subjects to wear or carry them, and devices that require electrodes to be attached to the body, which may cause discomfort.

To this end, we introduce a new emotion recognition approach using footstep-induced floor vibration signals. The main intuition of this approach is that people's gait patterns change under various emotions, thus inducing distinct structural vibration patterns as they walk. Compared to other methods, our approach is non-intrusive, insensitive to visual obstructions, and has fewer privacy concerns. Previous works have shown that different emotional states can lead to variations in walking speed, ankle and knee rotation, and ground reaction force [17]. These variations in gait are reflected in the footstep-induced floor vibration signals, which provide insights into the pedestrian's emotional state. By analyzing the footstep-induced floor vibration signals, we can capture the pedestrian's gait patterns and thus infer their emotional states.

The main research challenge in developing our approach is that emotions have both explicit and implicit effects on gait, making the explicit gait parameters insufficient to describe such a complex relationship. The explicit effects of the intervention can be characterized by analyzing gait parameters such as step frequency, double stance time, and other relevant metrics. The implicit effects on the gait pattern may not be fully described in gait parameters, requiring more implicit features. To overcome these challenges, our approach integrates various features, including both gait parameter features (step frequency, double support time, etc.) and signal-based features (Fourier transform coefficients, energy contour, etc.) extracted from the footstep-induced vibration signals to obtain better emotion recognition results. Signal-based features focus on understanding the intrinsic properties of the signal. Gait parameter features capture information related to the pedestrian's walking characteristics. Our approach is evaluated in a real-world walking experiment involving 5 participants with over 100 minutes of footstep-induced floor vibration signals. Our results show that our approach achieves a mean absolute error of 1.33 for valence score estimation and 1.26 for arousal score estimation out of a score range from 1 to 9, and accuracy of 72% for High / Low valence classification and 82% for High / Low arousal classification, which is comparable to other more intrusive state-of-the-art gait-based emotion recognition methods (range around 60% to 80% [18–20]).

The main contributions of this paper are:

1. We propose a novel approach for recognizing emotions based on footstep-induced floor vibration signals. To the best of our knowledge, this is the first emotion recognition system based on footstep-induced vibration signals.
2. We develop and integrate both gait parameter features (step frequency, double support time, etc.) and signal-based features (Fourier transform coefficients, energy contour, etc.) to capture both the explicit and implicit effects of emotions on gait.
3. We conduct a real-world walking experiment with 5 participants, which consists of over 100 minutes of walking samples, and demonstrate the effectiveness of our approach.

The rest of the paper presents the background information, our emotion recognition system, evaluation with a real-world walking experiment, and conclusions.

BACKGROUND: EMOTION, FOOTSTEP-INDUCED FLOOR VIBRATION

The main intuition of our approach is that emotions influence physiological activities, and thus affect gait patterns, ultimately leading to footstep-induced floor vibrations through the interaction between the foot and the floor (See Fig. 1). This section includes an introduction to the description of emotions, then the relationship between emotions and gait patterns, and an explanation of the influence of gait patterns on footstep-induced floor vibrations.

Figure 1. Relationship between emotion, gait, and footstep-induced floor vibration

In this paper, we use the emotional states described by Russel's 2D circumplex model of emotion (See Fig. 2), which is one of the most widely utilized models for describing emotion [21]. Russell's model conceptualizes emotion as comprising two independent dimensions: valence and arousal. Each quadrant of the 2D space represents a specific type of emotion.

Emotional valence and arousal correspond to the physiological activity associated with each emotion, thus affecting gait patterns (See Fig. 1). High arousal emotions like anger and anxiety are associated with significant increases in physiological activity (e.g., heart rate) as compared to low physiological emotions like sadness and reflection [22]. Similarly, emotional valence refers to the "pleasantness" of the emotion being experienced. Emotions like happiness and joy are "pleasant" emotions whereas emotions like sadness and anger are "unpleasant emotions". Past research suggests that both emotional valence and arousal are independently associated with gait patterns. For instance,

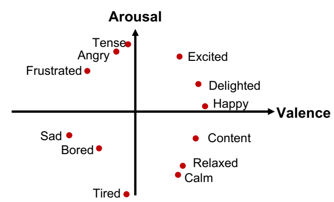


Figure 2. Russel's circumplex model of emotions

some research suggests that elicitation of pleasant vs unpleasant emotions results in systematically different patterns of gait [23]. Similarly, other research suggests that the valence component of emotions can be predicted independently using features derived from gait [24].

Different gait patterns can result in distinct patterns of footstep-induced vibration signals due to the interaction of the forces between the foot and the structure (See Fig. 1). Previous studies have demonstrated that footstep-induced floor vibration signals carry valuable information about human gait patterns, which can be used for a variety of applications, such as occupant identification [25, 26], localization [27, 28], activity recognition [29–32], and health status estimation [33]. The time difference between detected footstep signals can indicate step frequency. Additionally, a higher foot lift can result in a stronger force applied to the floor, leading to larger energy in the vibration signals. The contact type can also impact the time-frequency spectrum of the vibration signals [33]. For example, dragging feet can cause friction with the floor, resulting in higher frequency signals. By extracting features from the footstep-induced vibration signals, we can gather information about their gait patterns, allowing us to infer their emotional states.

EMOTION RECOGNITION SYSTEM USING FOOTSTEP-INDUCED FLOOR VIBRATION

Our emotion recognition system mainly consists of three modules (See Fig. 3): 1) Footstep Detection and Data Preprocessing, 2) Emotion-Related Feature Extraction, and 3) Emotion Recognition Based on Combined Emotion-Related Features. In the first module, the footstep-induced floor vibration signal sequence is segmented into individual footstep signals and pre-processed. In the second module, we extract both gait parameters and signal-based features from the footstep-induced floor vibration signals to capture the gait pattern. In the third module, the extracted features are embedded using different types of layers and then fed into a multilayer perceptron model to estimate the pedestrians' emotional states.

Module 1: Footstep Detection and Data Preprocessing. In Module 1, the footstep-induced floor vibration signals are captured and segmented into individual footstep signals and then pre-processed for further analysis. The footstep-induced floor vibrations are captured and converted into electrical signals and amplified to increase the signal-to-noise ratio. We develop an algorithm to detect footstep signals and separate them from the whole vibration signal sequence. To achieve this, we detect the major peaks of the absolute value of the wavelet coefficient of the vibration signals over time domain [28],

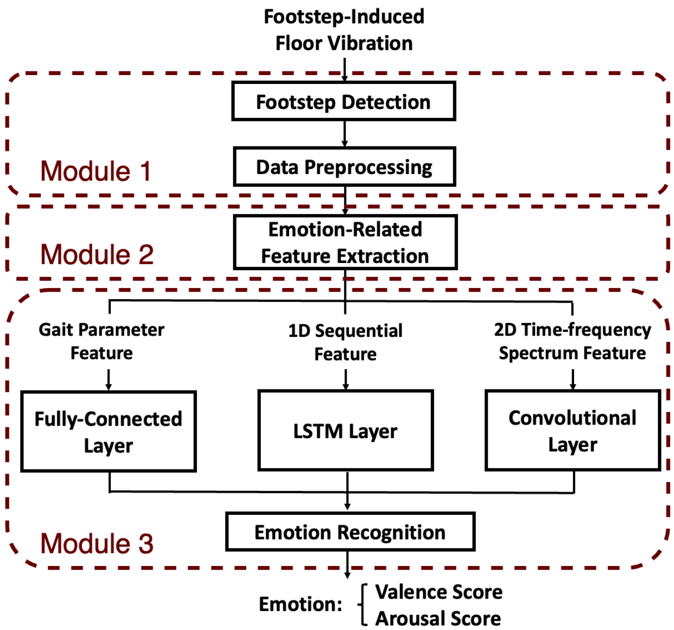


Figure 3. System Overview

which is caused by the force applied to the floor when walking. Assuming that each footstep last around 0.35 seconds (average footstep signal duration time based on our observation), the 0.35-second sequence window centered around the detected peak is separated as a footstep signal sequence. Due to hardware limitations, there may be instances of signal clipping, which can affect the signal shape and later feature extraction. To address this issue, we first detect the clipped sections of the signal, where the signal stays at the upper or lower limit range for an extended period. Then we use the sample points around the clipping section to perform polynomial interpolation, thereby reducing the clipping effect on the signal.

Module 2: Emotion-Related Feature Extraction. To better capture the intricate features of gait patterns related to emotions, we extract both gait parameter features [1, 17], as well as commonly used signal-based features [34] based on the footstep-induced vibration signals. The gait parameter features include step frequency, double support time, peak ratio of the heel-strike and toe-off, and footstep friction indicator. Previous studies have shown that during high-arousal emotions, the walking speed is faster and the joint rotation angle is larger, with larger rotation angles of knees observed during happy emotions [1]. Hip rotation is affected by leaning backward under low arousal emotions [1]. Step frequency can indicate the speed of walking. Double support time and peak height ratio reflect the pedestrian's center of gravity leaning forward or backward, which indicates the pedestrian's relaxation level. The footstep friction is usually caused by the low height of the foot lift during the swing phase.When the foot is moving forward, the reduced height of the foot lift results in the foot during the swing phase remaining partially in contact with the ground, causing friction to occur. The step frequency is calculated based on the time difference between the current step and the next consecutive detected footstep. The double support time is calculated as the time difference between the heel-strike time and the toe-off time [35]. To detect heel strike and toe-off, we apply wavelet transform to the signal and identify the peak of the higher frequency band (>=100 Hz) and lower frequency band (<100 Hz), respectively. Friction is detected through the higher frequency bands (>=100 Hz), as it usually generates shorter-duration signals with higher frequency components. We use an indicator variable to represent the presence (1) or absence (0) of friction during each footstep.

The signal-based features are the smoothed signal energy contour (0.01-second window), Fourier transform spectrum, and continuous wavelet transform spectrum. Compared to gait parameter features, the signal-based features with higher dimensions provide a representation of the footstep-induced vibration signals with less loss of original information, allowing us to extract additional information related to the pedestrian's gait and emotional states. The energy contour of a vibration signal corresponds to the strength applied to the floor by the feet. The Fourier transform spectrum provides frequency domain information of the signal, which indicates the pedestrian's gait pattern. For example, when the ankle rotation is larger, the heel strike results in a sudden impulse to the floor, causing higher-frequency vibrations. Additionally, 2-dimensional time-frequency features, such as the continuous wavelet transform spectrum, can indicate both temporal and spatial information, allowing for the detection of frequency distribution changes over time.

By combining both gait parameters and signal-based features, we can better characterize gait patterns in quantitative ways, which allows us to capture the complex relationship between gait patterns and emotion.

Module 3: Emotion Recognition Based on Combined Emotion-Related Features. In Module 3, we develop various methods to extract emotional state information from features characterized by different types of data dependencies. Subsequently, a multilayer perceptron is utilized to estimate the valence and arousal scores of the pedestrians. The signal-based features can be grouped into 1-dimensional sequential features (energy contour (0.01-second window), Fourier transforms spectrum), and 2-dimensional time-frequency features (continuous wavelet transforms spectrum) based on the data dependency type for subsequent extraction of information with different layers. They are later processed with different layers to capture different types of data dependencies. We select long-short-term-memory (LSTM) layers to process the temporal sequential features, as LSTM is a type of recurrent neural network (RNN) that uses gate functions to track long-term dependencies in the sequences, making it well-suited for sequential data. In contrast, gait parameter features do not have long-term or short-term dependencies like sequence data, so we select fully connected layers. The time-frequency spectrum is a 2-dimensional feature that requires capturing both the time and frequency dependencies. Hence, we select a convolutional layer, which is proficient at extracting information from 2-dimensional data and thus capturing the time-frequency dependencies in the vibration signals. We concatenate the embedded features and feed them into a multilayer perceptron, which consists of two fully connected layers. The outputs of the model are the valence score and arousal score.

EVALUATION WITH REAL-WORLD WALKING EXPERIMENT

To evaluate our system, we conduct an experiment with five participants, with each

TABLE I. THE VALENCE AND AROUSAL SCORE VARIATIONS BETWEEN INDIVIDUALS AND WITHIN INDIVIDUALS AND THE AVERAGE PRE-POST DIFFERENCES IN VALENCE AND AROUSAL SCORES.

	Range of Valence	Range of Arousal	Average Pre-post Difference of Valence	Average Pre-post Difference of Arousal
Person 1	7	8	3.125	4.5
Person 2	1	4	0.625	1.875
Person 3	2	4	1	1.5
Person 4	5	4	2.25	1.875
Person 5	6	6	1.75	2.375

participant walking back and forth for nine trials of 2-3 minutes each on a wooden floor. As illustrated in Fig. 4, four geophone sensors (SM-24) are attached to the floor to capture the footstep-induced floor vibration signals, with a sampling rate of 500 Hz. The signal from the sensor is then amplified by a hardware amplifier to increase the signal-to-noise ratio. Prior to walking, participants are provided with specific music and lighting to elicit particular emotions. The music clips used are selected from the Previously-Used Musical Stimuli (PUMS) database [36], while the light strip's (Govee RGBIC LED Strip Lights) color is chosen as warm and bright for positive emotional stimuli and shine brightly with high arousal emotional stimuli. In contrast, a cold color is used for negative emotional stimuli and remains still for low arousal emotional stimuli [37]. In total, eight sets of emotional stimuli were administered, with two each for high valence high arousal, high valence low arousal, low valence high arousal, and low valence low arousal. The order of the stimulus sets is randomized for each participant. For each trial, participants are asked to walk on the platform back and forth for about 2-3 minutes. All experiments are conducted in accordance with the approved IRBs.

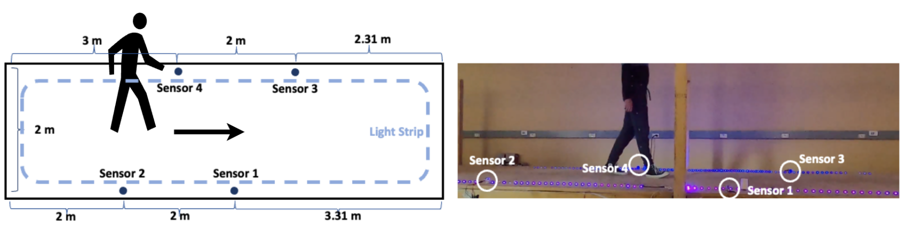
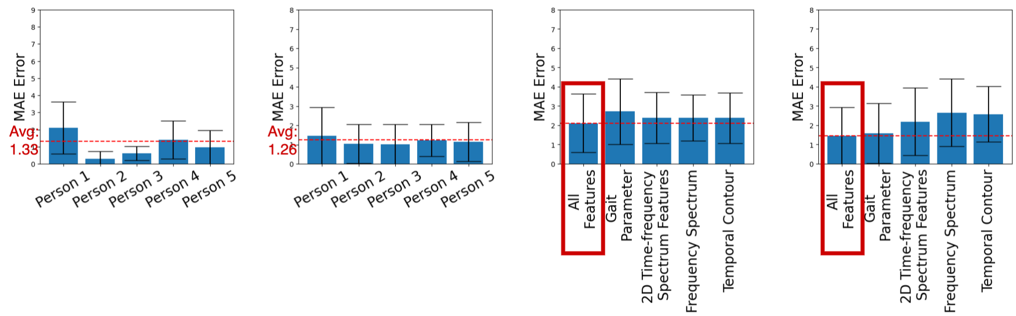


Figure 4. Experiment Setup

The ground truth labels in our evaluation are the self-reported valence and arousal scores. After each 2-3 minute walking trial, participants are asked to complete the Self-Assessment Manikin (SAM) survey scale [38], which involves rating their subjective valence and arousal levels on a scale of 1 to 9.

The effectiveness of the emotional stimulus in the experiment is validated through the variation in valence and arousal scores, both between individuals and within individuals, as well as the average pre-post differences in these scores (See Table. I). Greater average pre-post differences in valence and arousal scores indicate a stronger impact of emotion stimuli on participants. It is observed that the emotional stimuli have a notably strong impact on Person 1, a moderately strong effect on Persons 4 and 5, and a minor effect on Persons 2 and 3.

Our method achieves promising results on the dataset of five participants, comprising 12918 footstep signal samples. Specifically, the mean absolute errors of the valence and arousal scores are 1.33 and 1.26, respectively, with a score range of 1 to 9. Figs. 5a and 5b present the estimation results for the five participants, highlighting the error of our model in estimating the emotional states of pedestrians.



(a) (b) (c) (d)

Figure 5. Evaluation Result showing the mean absolute error (MAE) of (a) valence score for 5 participants (Average MAE value: 1.33) (b) arousal score for 5 participants (Average MAE: 1.26) (c) valence score comparison: Combining Features vs. Single Type Feature (d) arousal score comparison: Combining Features vs. Single Type Feature

For the emotion classification task, emotions are classified as high or low valence (arousal) with the threshold set as the average valence (arousal) score for each pedestrian. We achieve a classification accuracy of 72% for high/low valence classification and 82% for high/low arousal classification, which is comparable to other more intrusive state-of-the-art gait-based emotion recognition methods (range around 60% to 80% [18–20]).

Moreover, we compare the performance of our model when using a combination of features and when using only one type of feature, using the data from person 1. As shown in Fig. 5c and Fig. 5d, the results indicate that our proposed method outperforms the use of a single feature type in terms of estimation error, demonstrating its effectiveness.

CONCLUSION

In this paper, we introduce a novel approach for emotion recognition using footstep-induced floor vibration signals. The main challenge in developing our approach is that emotions have both explicit and implicit effects on gait, making the explicit gait parameters insufficient to describe such a complex relationship. To this end, we develop a set of emotion-sensitive features from the vibration signals, including gait parameters, sequential features, and time-frequency spectrum features to capture both the explicit and implicit effects of emotion on gait. These features are then combined and fed into a multilayer perceptron for emotion estimation. Our approach achieves promising results on a dataset of 5 participants, with a mean absolute error of 1.33 for valence score estimation and 1.26 for arousal score estimation over 100 minutes of walking out of an overall score range of 1 to 9. For the emotion classification task, emotions are classified as high or low valence (arousal) with the threshold set as the average valence (arousal) score of the pedestrian. We are able to achieve an accuracy of 72% for the High / Low valence classification and 82% for the High / Low arousal classification in the emotion classification task, which is comparable to other more intrusive state-of-the-art gait-based emotion

17. Xu, S., J. Fang, X. Hu, E. Ngai, W. Wang, Y. Guo, and V. C. Leung. 2022. "Emotion recognition from gait analyses: Current research and future directions," *IEEE Transactions on Computational Social Systems*.
18. Quiroz, J. C., E. Geangu, and M. H. Yong. 2018. "Emotion recognition using smart watch sensor data: Mixed-design study," *JMIR mental health*, 5(3):e10153.
19. Chiu, M., J. Shu, and P. Hui. 2018. "Emotion recognition through gait on mobile devices," in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, IEEE, pp. 800–805.
20. Ahmed, F., B. Sieu, and M. L. Gavrilova. 2018. "Score and rank-level fusion for emotion recognition using genetic algorithm," in *2018 IEEE 17th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)*, IEEE, pp. 46–53.
21. Russell, J. A. 1980. "A circumplex model of affect." *Journal of personality and social psychology*, 39(6):1161.
22. Appelhans, B. M. and L. J. Luecken. 2006. "Heart rate variability as an index of regulated emotional responding," *Review of general psychology*, 10(3):229–240.
23. Naugle, K. M., C. J. Hass, J. Joyner, S. A. Coombes, and C. M. Janelle. 2011. "Emotional state affects the initiation of forward gait." *Emotion*, 11(2):267.
24. Randhavane, T., U. Bhattacharya, K. Kapsaskis, K. Gray, A. Bera, and D. Manocha. 2019. "Identifying emotions from walking using affective and deep features," *arXiv preprint arXiv:1906.11884*.
25. Pan, S., T. Yu, M. Mirshekari, J. Fagert, A. Bonde, O. J. Mengshoel, H. Y. Noh, and P. Zhang. 2017. "Footprintid: Indoor pedestrian identification through ambient structural vibration sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):1–31.
26. Dong, Y., J. Fagert, P. Zhang, and H. Y. Noh. 2022. "Stranger Detection and Occupant Identification Using Structural Vibrations," in *European Workshop on Structural Health Monitoring: EWSHM 2022-Volume 1*, Springer, pp. 905–914.
27. Mirshekari, M., S. Pan, J. Fagert, E. M. Schooler, P. Zhang, and H. Y. Noh. 2018. "Occupant localization using footstep-induced structural vibration," *Mechanical Systems and Signal Processing*, 112:77–97.
28. Dong, Y., J. Liu, and H. Y. Noh. 2022. "GaitVibe+: Enhancing Structural Vibration-based Footstep Localization Using Temporary Cameras for In-home Gait Analysis," *arXiv preprint arXiv:2212.03377*.
29. Pan, S., M. Berges, J. Rodakowski, P. Zhang, and H. Y. Noh. 2019. "Fine-grained recognition of activities of daily living through structural vibration and electrical sensing," in *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pp. 149–158.
30. Bonde, A., S. Pan, M. Mirshekari, C. Ruiz, H. Y. Noh, and P. Zhang. 2020. "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)*, pp. 216–222, doi:10.1109/IoTDI49375.2020.00028.
31. Fagert, J., A. Bonde, S. Srinidhi, S. Hamilton, P. Zhang, and H. Y. Noh. 2022. "Clean vibes: Hand washing monitoring using structural vibration sensing," *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(3):1–25.
32. Ruiz, C., S. Pan, A. Bannis, M.-P. Chang, H. Y. Noh, and P. Zhang. 2020. "IDIoT: Towards Ubiquitous Identification of IoT Devices through Visual and Inertial Orientation Matching During Human Activity," in *2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pp. 40–52, doi:10.1109/IoTDI49375.2020.00012.

33. Dong, Y., J. J. Zou, J. Liu, J. Fagert, M. Mirshekari, L. Lowes, M. Iammarino, P. Zhang, and H. Y. Noh. 2020. “MD-Vibe: physics-informed analysis of patient-induced structural vibration data for monitoring gait health in individuals with muscular dystrophy,” in *Adjunct proceedings of the 2020 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2020 ACM international symposium on wearable computers*, pp. 525–531.
34. Giannakopoulos, P. and A. Pikrakis. 2014. *Introduction to audio analysis: a MATLAB® approach*, Academic Press.
35. Dong, Y. and H. Y. Noh. “Structure-Agnostic Gait Cycle Segmentation for In-Home Gait Health Monitoring through Footstep-Induced Structural Vibrations,” .
36. Warrenburg, L. 2021. “The PUMS database: A corpus of previously-used musical stimuli in 306 studies of music and emotion,” *Empirical Musicology Review*, 16(1):145–150.
37. Xie, X., J. Cai, H. Fang, X. Tang, and T. Yamanaka. 2022. “Effects of colored lights on an individual’s affective impressions in the observation process,” *Frontiers in Psychology*, 13:938636.
38. Bradley, M. M. and P. J. Lang. 1994. “Measuring emotion: the self-assessment manikin and the semantic differential,” *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59.