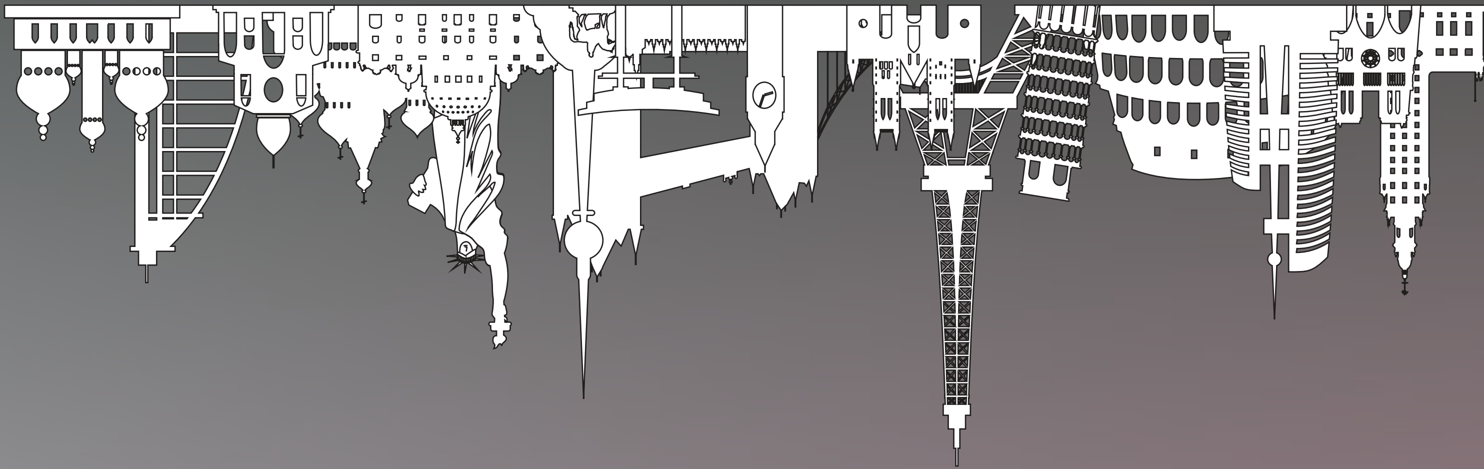
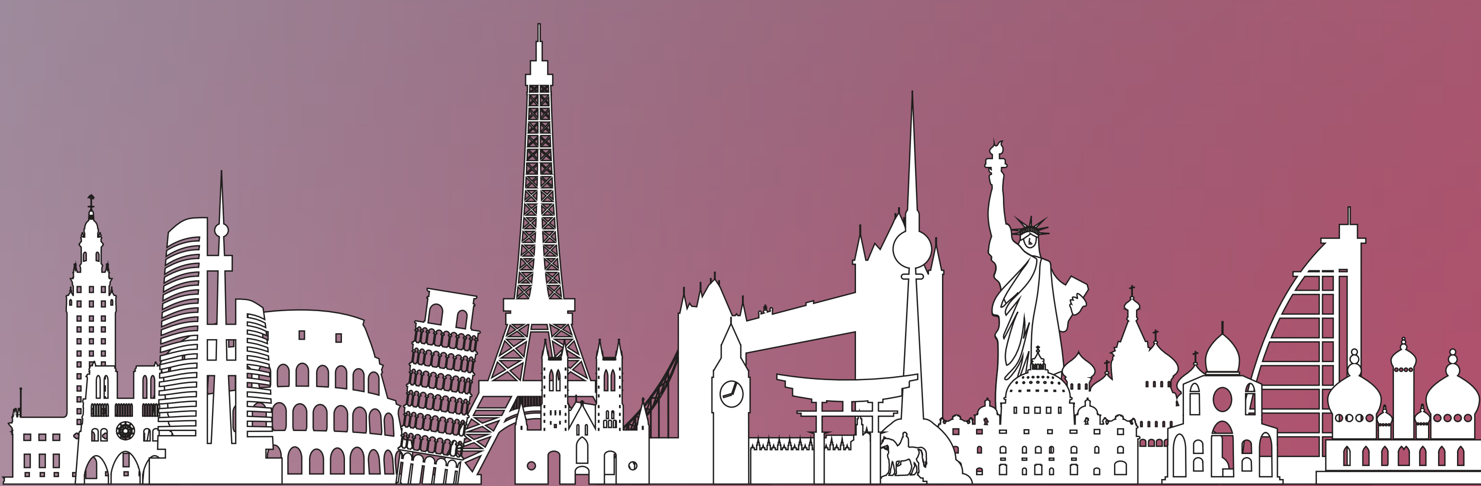


EarComp 2021

2nd International Workshop on Earable Computing

In conjunction with UbiComp 2021

September 25th, 2021, All Over the World



Detecting Verbal and Non-Verbal Gestures Using Earables

Matias Laporte
matias.laporte@usi.ch
Università della Svizzera italiana (USI)
Switzerland

Preety Baglat
preety.baglat@usi.ch
Università della Svizzera italiana (USI)
Lugano, Switzerland

Shkurta Gashi
shkurta.gashi@usi.ch
Università della Svizzera italiana (USI)
Lugano, Switzerland

Martin Gjoreski
martin.gjoreski@usi.ch
Università della Svizzera italiana (USI)
Lugano, Switzerland

Silvia Santini
silvia.santini@usi.ch
Università della Svizzera italiana (USI)
Lugano, Switzerland

Marc Langheinrich
marc.langheinrich@usi.ch
Università della Svizzera italiana (USI)
Lugano, Switzerland

ABSTRACT

Verbal and non-verbal activities convey insightful information about people's affect, empathy, and engagement during social interactions. In this paper, we investigate the usage of inertial sensors to recognize verbal (e.g., *speaking*), non-verbal (e.g., *head nodding*, *shaking*) and other activities (e.g., *eating*, *no movement*). We implement an end-to-end deep neural network to distinguish among these activities. We then explore the generalizability of the approach in three scenarios: (1) using new data to detect a known activity from a known user, (2) detecting a novel activity of a known user and (3) detecting the activity of an unknown user. Results show that using accelerometer and gyroscope sensors, the model achieves a balanced accuracy of 55% when tested on data from a new user, 41% on a new activity of an existing user, and 80% on new data of a known activity from an existing user. The results are between 7-47 percentage points higher than baseline classifiers.

CCS CONCEPTS

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

KEYWORDS

Datasets; Earable Computing; Head Gestures Recognition; Memory Recall

ACM Reference Format:
Matias Laporte, Preety Baglat, Shkurta Gashi, Martin Gjoreski, Silvia Santini, and Marc Langheinrich. 2021. Detecting Verbal and Non-Verbal Gestures Using Earables. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3460418.3479322

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

ACM ISBN 978-1-4503-8461-2/21/09…$15.00
https://doi.org/10.1145/3460418.3479322

1 INTRODUCTION

Our motivation for detecting verbal and non-verbal activities is rooted in our work on building human memory augmentation systems. Using earable computing, we attempt to recognize different types of human activities, in particular head gestures, with the purpose of detecting when a social interaction is taking place. This is because the presence of others and our interactions with them play important roles in our memories, both during the formation of memory [32] and at retrieval time [6]: moments of social interactions might be easier to remember (formation time), and remembering a particular interaction might also help to remind us of particular details (retrieval time).

The importance of human memory for our daily lives cannot be overstated. It gives us an identity, lets us remember future intentions, carry quotidian tasks, and obtain new knowledge. It also allows us to share experiences and maintain and nurture relationships [20]. Therefore, civilization has applied increasingly complex methods to preserve its memories and overcome their failures. Today, capture technology such as cameras, voice recorders, and fitness trackers are coming close to making total capture (and, consequently, total recall [2]) a possibility, if not already a reality [13]. However, even if every part of our lives is captured and recorded, it is far from trivial to then use this information to aid our memory.

Memory augmentation systems will only succeed as long as they are able to appropriately select the relevant memories for the user [30]. In fact, instead of presenting the user a fully recorded memory, these systems should take advantage of the power of *memory cues* – objects or events that help us remember our original memory or intent. By prompting the user with such a (small) cue, they will be able to recall the original experience in great detail. One key challenge here is to identify appropriate cues among the recorded data that, when played back to the user, will trigger such recall. Social interaction might mark important moments that may make useful memory cues.

Social interaction can be easily detected using audio sensing: detecting a conversation is a sure sign of interpersonal activity. Similarly, closely tracking the movements and orientation of people could allow us to identify social interaction. Alternatively, a wearable camera may pick up faces of others and identify social interaction. All three options rely on highly sensitive personal data. Instead, we seek to identify head gestures to detect both verbal and nonverbal social interaction from inertial signals.

To summarize, this paper presents the following contributions:

- We present a new dataset comprised of accelerometer and gyroscope data collected using ear-worn devices. The dataset was collected from 10 participants performing 5 scripted activities: *nodding*, *speaking*, *eating*, *staying*, and *head shaking*.
- We investigate the feasibility of using a deep Convolutional Neural Network (CNN) to recognize activities related to social interactions such as verbal (e.g., *speaking*) and non-verbal (e.g., *nodding*) gestures, as well as gestures unrelated to interactions (e.g., *eating*).

2 RELATED WORK

The continuous development of unobtrusive wearable sensors has made possible the recording of new types of data in uncontrolled settings. Of particular interest to our work is the use of earable sensors, i.e., head-mounted in-ear/behind-the-ear sensors, to detect speech and head gestures, as cues for human interaction. As previously mentioned, other approaches (e.g., cameras and microphones) require an involved setup with additional privacy issues to consider.

Current earable devices can accommodate several sensors (e.g. accelerometers, gyroscopes, microphones or biometric sensors) and actuators (e.g., speaker) in a comfortable size with decent battery autonomy, allowing not only sound and head movement measurements, but also of head rotation and bio-signals, among others.

2.1 Earable Systems

Earable sensors have been proposed as a tool with "enormous potential in accelerating our understanding of a wide range of human activities in a nonintrusive manner", with applications ranging from "health tracking" to "contextual notification management" , including "cognitive assistance" and "lifelogging" [22].

Among the applications to deepen our understanding of human behaviour, Frohn et al. [11] have used an earable sensor to characterize the emotional intent of study participants performing a series of scripted scenes. Although the results were limited due to the reduced sample size and the use of non-actors, they showed that participants act more energetically and in sync when the scenes have a positive intent, than otherwise.

Röddiger et al. [29] instead used in-ear accelerometer and gyroscope sensors for health tracking, by measuring the respiration rate of the participants.

Other applications include: EarDynamic [33], a biometric-based authentication method which models users' ear canal deformation through the emission of inaudible audio signals and their reflections; and EarBuddy [34], a gesture recognition system which uses the microphones on the earbuds to detect different types of finger touches in the face.

Although these approaches have done novel applications with the available earable technology, none of them have focused on recognizing verbal and non-verbal activities from inertial signals.

2.2 Human Behaviour Detection

Earable sensors benefit particularly from the proximity and contact with the face to be able to distinguish the movements of the jawbone and the activation of the different muscles.

EarBit [1], for example, was a prototype with multimodal (acoustic, motion) sensors to detect chewing episodes. It used an optical proximity sensor to measure the deformation of the ear canal produced by the movement activity of the mandibular bone, and a 9-axis Inertial Measurement Unit (IMU) to capture the movement of the temporalis muscle, used when chewing. EarBit also included a microphone located around the neck to detect swallowing events. A chest-mounted GoPro was used as ground truth collector. Auracle [3] is another example of eating detection, but with the use of a contact microphone instead, and an unobtrusive ground truth collector embedded into a cap.

In STEAR [28], ear mounted IMU sensors have been proposed as a new approach for step counting, with the benefit of not being affected by random motions of leg and hand, like it would happen with a smartphone or a smartwatch, respectively.

Only a few researchers focused on the recognition of head gestures and human activities, even in social interaction settings, with the use of earables.

Gjoreski et al. [16] used a 9-axis IMU to detect 8 individual daily life activities from a dataset of 4 subjects. Ferlini et al. [10] used an ear-worn device to track head rotations while performing activities like chewing and speaking. Min et al. [26] used an IMU sensor and a microphone for monitoring conversational well-being, using models that recognize speaking activities, altogether with stress and emotion detection. Tan et al. [31] used earable devices to detect the head orientation of interacting groups and used it as a cue for directed social attention. Lee et al. [24] focused instead on the recognition of smile and frowns gestures, while Islam et al. [21] proposed an activity recognition framework differentiating between head and mouth related activities (e.g. head shaking, nodding, eating and speaking), and normal activities (e.g. staying, walking and speaking while walking).

Our work further expands on Islam et al.'s by considering the detection of verbal and non-verbal gestures in the context of social interactions, with the intent of marking part of those moments as important for the use of memory cues.

2.3 Human Memory Augmentation

The idea of a system that stores one's digital records (e.g., documents, images, multimedia etc.) for a lifetime goes back to the 1945 vision of the Memex by Vannevar Bush [7]. While Bush did not detail the exact technology for implementing his vision, he predicted an era when storage will be virtually unlimited. Some 60 years later, the MyLifeBits project attempted to fulfill the promise of Bush's vision [14]. MyLifeBits started as a platform that could log all personal information generated and accessed on a PC, but its memory enhancing aspects quickly emerged [13]. More recently, Davies et al. [9] described the vision and core architectural building blocks of a future pervasive memory augmentation ecosystem, while Harvey et al. described the role of lifelogging technology in this vision [19].

3 DATA COLLECTION

We provide below details about the participants, the type of data we collected, the tools used to do it and the data collection procedure.

3.1 Collected Data and Tools

For each participant, we recorded data from accelerometer and gyroscope sensors. To collect sensor data, we used the eSense earbuds

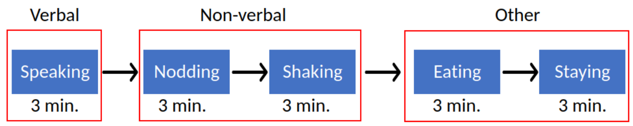


Figure 1: Experimental Protocol. Participants performed each activity for 3 minutes, with no particular order.

developed by Kawsar et al. [23] at Nokia Bell Labs. The eSense earbuds are equipped with 6-axis Inertial Measurement Unit (IMU) sensors, comprised of 3-axis accelerometer and 3-axis gyroscope sensors [5]. Being worn on the ear, the eSense is suited for gathering sensor data for detecting human gestures in an unobtrusive and continuous manner.

The acccelerometer sensor measures the acceleration of the device in G-force [5]. The gyroscope sensor measures the rotation of the device in degrees per second (deg/sec) [5]. Acceleration and gyroscope data measured from ear-worn devices have been shown to reflect the movements of the head and facial muscles [1, 22, 24]. Thereby, they seem suitable to detect whether a person is interacting with another individual or not. The eSense device also contains a microphone sensor, which could be used to detect verbal activities during social interactions such as e.g., speaking. However, microphone use raises privacy concerns for users [25, 26] and is not suitable to detect non-verbal types of interactions (e.g., nodding).

To collect the sensor data, we use the eSense app[1]. The eSense app is a smartphone application developed for the Android operating system. The application was initially implemented by Islam et al. [21] and then extended by Frohn et al. [11]. The app connects via Bluetooth Low Energy (BLE) to the eSense earbuds and obtains the sensor data. We set the sampling rate of the sensors to 25 Hz.

3.2 Participants and Procedure

We recruited 10 participants (6 females and 4 males). The majority of the participants were between 18 - 34 years old, and one was over 55 years old. Participants had different occupations such as e.g., worker (3), postdoctoral researcher (1), Ph.D. student (1), and University student (5).

Before meeting each participant, we charged the eSense and the mobile phone. Previous to the experiment, the researcher responsible for running the data collection explained the study goal and the data collection procedure to the participant. All participants signed an informed consent form. The experimenter provided the left earbud to the participant and instructed how to wear it. The left eSense earbud was then connected to the Android application. The participant was then instructed to first select the activity they wished to perform, and then to select the start button on the app to record the sensor data. At the end of recording an activity, the participants stopped the data recording and repeated the same procedure for another activity.

The participants performed five scripted activities, namely, *nodding*, *speaking*, *eating*, *standing still*, and *head shaking*. We choose these activities to investigate whether verbal and non-verbal interaction activities (e.g., *speaking* and *nodding*) are distinguishable from other head and mouth-related activities (e.g., *eating*, *head shaking*) as well as no activity at all (e.g., *standing still*). The participants performed each activity for 3 minutes, one after the other, and they were free to pick the order in which the activities were performed. A simple diagram of this procedure can be seen in Figure 1.

4 DATA ANALYSIS

The main goal of our work is to develop a method to recognize human verbal, non-verbal interactions or no interactions using inertial signals. In this section, we describe the end-to-end deep learning pipeline we developed as well as the evaluation procedures, metrics and baselines used.

4.1 Data Pre-processing

To pre-process the signals, we follow common pre-processing steps used in the literature for human activity recognition from inertial signals [5]. In particular, after dividing the dataset into train and test splits, we segment the sensor data for each split into 4 seconds windows with 75% overlap. After the segmentation, our final dataset contains 1210 *speaking* samples, 1162 *nodding*, 1272 *eating*, 1179 *head shaking* and 1127 *standing still*. The measurement unit of acceleration data is converted to ± 4g and gyroscope data to ± 500 deg/s directly in the application used to collect the data.

4.2 Convolutional Neural Network (CNN)

We developed an end-to-end CNN, which takes as input the 4-second windows of raw 3-axis accelerometer and gyroscope signals (see Figure 2). The accelerometer and gyroscope sensor data is first processed by three convolutional layers, each with a kernel size of 7, 128 feature maps and ReLU activation function. These layers learn feature representation from the raw sensor data. The output of the last convolutional layer is then flattened and provided as input to a max pooling layer. To avoid over-fitting, we employed dropout regularization with a dropout rate of 0.5. The output of the last layer of the model is provided as input to a `sigmoid` function, which returns a k dimensional output with estimated probability between 0 and 1, where k is the number of activity classes, which is 3 (*non-verbal*, *verbal*, or *other*).

4.3 Evaluation Procedures

To evaluate the performance of the CNN classifier, we follow common procedures in machine learning [15, 27]. In particular, we investigate three validation procedures described as following.

Leave-one-part-out (LOPO) validation procedure uses the data of all participants, except one, and the first 80% of the data of each activity from the left-out user in the training set. The remaining 20% of each activity of the left-out user is used in the test set. The procedure is repeated for all participants and the results are reported as average of all iterations. This approach verifies the ability of the model to generalize to unseen data of a known user. It also avoids the *temporal leak* issue, discussed in [8], which refers to situations when a model is trained on data from the future. With this approach, we ensure that the test set is posterior to the data in the training set.

Leave-one-activity-out (LOAO) evaluation approach uses all the data of all users, except one activity of one user, in the training set. The left-out activity of the user is kept as the test set. The same

[1]https://github.com/SabrinaFrohn/Esense

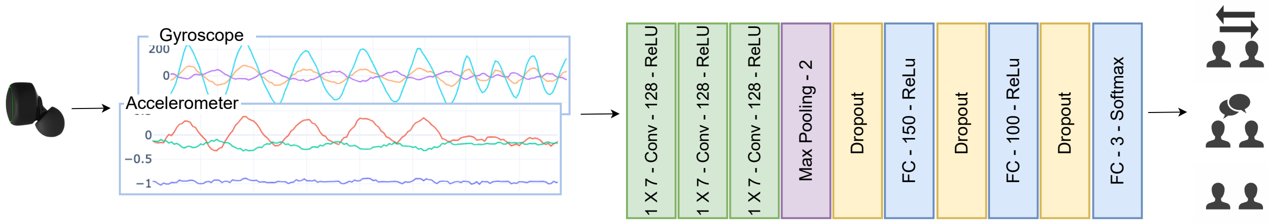


Figure 2: Overview of the end-to-end deep learning pipeline. The raw acceleremeter and gyroscope data collected from eSense earbuds is provided as input the CNN. The model classifies each data sample as a verbal, non-verbal or other activity.

procedure is repeated for all activities of the left-out user and then for all the users. We report the classification results as the average of all iterations. The main goal of this technique is to avoid having segments from a same trace (e.g., activity) collected from one user in both training and testing sets. This is because adjacent segments are not statistically independent, as discussed in [18]. This approach verifies the ability of the model to recognize new activities from a known user by learning from the presence of an activity in the training set from other users.

Leave-one-subject-out (LOSO) validation scheme uses the data of all users except one in the training set and the left-out user as the test set, as used in [24, 25]. This procedure is repeated for all the users. We report the classification results as the mean metrics for all users. This validation procedure ensures that the training set does not contain all of the activities from one user. With this technique, we aim to investigate the generalization of the model to new users.

4.4 Evaluation Metrics

To evaluate the performance of the model, we use *accuracy*, *balanced accuracy* and *F1*. Accuracy quantifies the number of samples correctly classified by the model [27]. Balanced accuracy score is defined as the average of recall score obtained in each class [4]. This score is suitable to compare the performance of imbalanced datasets because it also takes into consideration the minority class. To further explore the performance of the classifier in all the classes, we also report the weighted F1 metric. The F1 score is the harmonic mean of precision and recall [27].

4.5 Baseline Classifiers

We compare the performance of the CNN with Random Guess (`RG`) and Biased Random Guess (`BRG`) baselines. RG provides a classification uniformly at random. BRG takes into consideration the distribution of the classes in the training set and generates a biased prediction. In particular, BRG always predicts the most frequent label in the training set, as used in [12].

5 RESULTS

In what follows, we present and discuss the evaluation results. We first report the performance of the CNN using the different evaluation procedures and baseline classifiers described in Section 4. We then investigate the performance of each sensor separately (`unimodal`) and their combination (`multimodal`).

5.1 Evaluation Procedures Comparison

We first compare the performance of the CNN model using `LOPO`, `LOAO` and `LOSO` validation techniques. Figure 3 shows the balanced accuracy of CNN and baseline classifiers for each evaluation technique. In all cases, the results are higher than baseline classifiers. These results imply that it is feasible to use ear-worn devices to distinguish between verbal, non-verbal and other activities performed during social interactions. Overall the classification results using `LOPO` are significantly higher than using `LOAO` or `LOSO`. In particular, the CNN has a balanced accuracy of 80%, which is 25 and 39 percentage points increment from `LOSO` and `LOAO` validation techniques. As expected, the presence of annotated data from the test user allows the model to achieve a higher performance. Therefore, future systems that aim to distinguish between verbal, nonverbal and other activities using earbuds, should first train the model with data from the user to avoid the cold start problem. The performance drop of the CNN when using the `LOSO` or `LOAO` techniques, suggests that such systems are difficult to generalize to the data of a new user or a new activity of a user. Given that `LOPO` validation procedure provides the best results, in the next experiments we present more detailed results for this validation procedure.

5.2 Comparison to Baseline Classifiers

Figure 4 shows different classification metrics for RG, BRG and CNN classifiers, using the best validation procedure explored in this work, `LOPO`. In particular, balanced accuracy for the CNN is 80%, 36% for the RG and 33% for the BRG. Our model shows 44 and 47 percentage points increment compared to RG and BRG classifiers.

Figure 5 shows the balanced accuracy for each participant using the LOPO validation approach and the CNN, as the best model available among those tested. We observe that the performance of the CNN for the majority of the users is higher than 60%, with the exception being users P01 and P03.

5.3 Unimodal vs Multimodal

In this set of results, we investigate the performance of training with single (`unimodal`) and multiple sensors (`multimodal`). Unimodal refers to experiments where only one sensor's data (e.g., accelerometer) is used as input to the CNN model. Multimodal refers to experiments using both accelerometer and gyroscope data. Figure 6 presents the balanced accuracy scores obtained for unimodal and multimodal approaches using `LOPO` validation. The balanced accuracy for the accelerometer data is 75%, for the gyroscope data

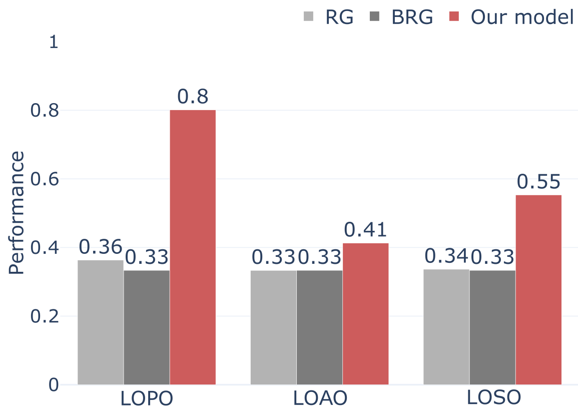


Figure 3: Results of CNN model using leave-one-part-out (LOPO), leave-one-activity-out (LOAO) and leave-one-subject-out (LOSO).

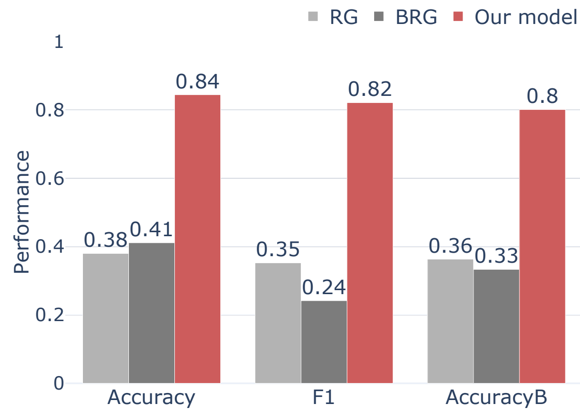


Figure 4: Accuracy, F1, and balanced accuracy (AccuracyB) for CNN, RG, and BRG classifiers using all signals as input and LOPO validation accuracy.

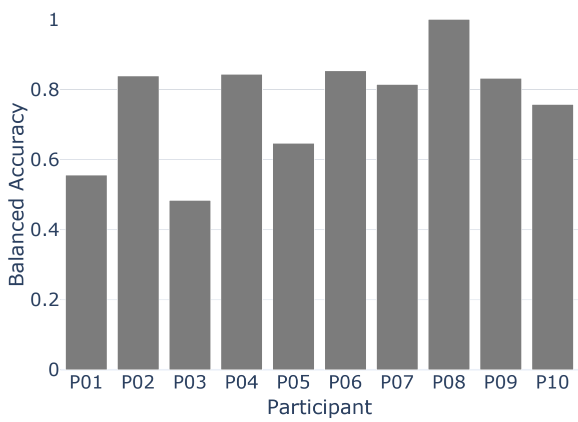


Figure 5: Balanced accuracy of CNN model for each participant using all signals as input and LOPO validation procedure.

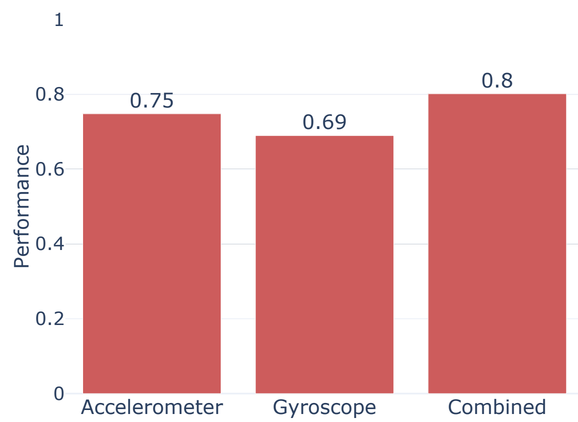


Figure 6: Comparison of the performance of the CNN model using only accelerometer, only gyroscope or both as input.

is 69%, and for their combination is 80%. We observe that the performance of the multimodal classifier is higher than the performance of unimodal classifiers by 5 and 11 percentage points respectively. These results imply that combining data from several sensors allows recognizing user's interactions better than using only one sensor. This outcome highlights the importance of considering not only the movements but also the rotation angle of the device during these activities, which is in line with other end-to-end deep learning studies on activity recognition [17].

6 LIMITATIONS AND FUTURE WORK

A limitation of our study is the data collected in a controlled setting. This might not reflect the challenges of collecting such data in real-world scenarios. In future work, we plan to run a larger study in naturalistic settings and verify the generalizability of our approach to new settings where users' movements are not constrained. In addition, we plan to investigate the relationship between the frequency of such activities and participants' memory recall.

While end-to-end deep learning offers the possibility to build activity recognition models without feature engineering, it also requires larger training datasets. As the dataset used in this study is relatively small, in the future we plan on implementing shallow, feature-based classifiers, and on increasing the size of the dataset to implement other end-to-end deep learning architectures.

In this work, we segmented accelerometer and gyroscope sensor data using a sliding window of 4 seconds. We plan on comparing different segmentation strategies (e.g., overlapping, non-overlapping) and window sizes, as in [25].

7 CONCLUSIONS

Social interaction presents an interesting feature for identifying moments that lend themselves to memory cue generation. Instead of using video, audio, or location tracking technology, we envision the use of unobtrusive inertial sensors to identify moments of social interaction – both verbal and non-verbal. This study showed that such sensors are useful for recognising a variety of head movements, which could be used as context information for future human memory augmentation systems.

ACKNOWLEDGMENTS

We thank Nokia Bell Labs for donating the eSense devices, and study participants for volunteering to be part of the experiment.

REFERENCES

[1] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. 2017. EarBit: Using Wearable Sensors to Detect Eating Episodes in Unconstrained Environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (PACM IMWUT)* 1, 3 (2017), 1–20.

[2] Gordon Bell and Jim Gemmell. 2009. Total Recall: How the e-Memory Revolution Will Change Everything. In *Total Recall: How the e-Memory Revolution Will Change Everything*. Penguin Books, London, United Kingdom.

[3] Shengjie Bi, Tao Wang, Nicole Tobias, Josephine Nordrum, Shang Wang, George Halvorsen, Sougata Sen, Ronald Peterson, Kofi Odame, Kelly Caine, et al. 2018. Auracle: Detecting Eating Episodes With an Ear-mounted Sensor. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (PACM IMWUT)* 2, 3 (2018), 1–27.

[4] Kay Henning Brodersen, Cheng Soon Ong, Klaas Enno Stephan, and Joachim M. Buhmann. 2010. The Balanced Accuracy and Its Posterior Distribution. In *2010 20th International Conference on Pattern Recognition*. IEEE, IEEE, Piscataway, NJ, United States, 3121–3124. https://doi.org/10.1109/ICPR.2010.764

[5] Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A Tutorial on Human Activity Recognition Using Body-Worn Inertial Sensors. *ACM Computing Surveys (CSUR)* 46, 3 (2014), 1–33.

[6] C. Burri. 1931. The Influence of an Audience upon Recall. *Journal of Educational Psychology* 22, 9 (1931), 683–690. https://doi.org/10.1037/h0070961

[7] Vannevar Bush et al. 1945. As we may think. *The atlantic monthly* 176, 1 (1945), 101–108.

[8] Francois Chollet. 2017. *Deep learning with Python*. Simon and Schuster, New York, NY, United States.

[9] Nigel Davies, Adrian Friday, Sarah Clinch, Corina Sas, Marc Langheinrich, Geoff Ward, and Albrecht Schmidt. 2015. Security and Privacy Implications of Pervasive Memory Augmentation. *IEEE Pervasive Computing* 14, 1 (2015), 44–53.

[10] Andrea Ferlini, Alessandro Montanari, Cecilia Mascolo, and Robert Harle. 2019. Head Motion Tracking through In-Ear Wearables. In *Proceedings of the 1st International Workshop on Earable Computing* (New York, NY, USA, 2019) *(EarComp'19)*. Association for Computing Machinery, New York, NY, USA, 8–13. https://doi.org/10.1145/3345615.3361131

[11] Sabrina A. L. Frohn, Jeevan S. Matharu, and Jamie A. Ward. 2020. Towards a Characterisation of Emotional Intent during Scripted Scenes Using In-Ear Movement Sensors. In *Proceedings of the 2020 International Symposium on Wearable Computers* (New York, NY, USA, 2020) *(ISWC '20)*. Association for Computing Machinery, New York, NY, USA, 37–39. https://doi.org/10.1145/3410531.3414292

[12] Shkurta Gashi, Elena Di Lascio, Bianca Stancu, Vedant Das Swain, Varun Mishra, Martin Gjoreski, and Silvia Santini. 2020. Detection of Artifacts in Ambulatory Electrodermal Activity Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–31.

[13] Jim Gemmell, Gordon Bell, and Roger Lueder. 2006. MyLifeBits: A Personal Database for Everything. *Commun. ACM* 49, 1 (Jan. 2006), 88–95. https://doi.org/10.1145/1107458.1107460

[14] Jim Gemmell, Gordon Bell, Roger Lueder, Steven Drucker, and Curtis Wong. 2002. MyLifeBits: Fulfilling the Memex Vision. In *Proceedings of the Tenth ACM International Conference on Multimedia* (Juan-les-Pins, France) *(MULTIMEDIA '02)*. Association for Computing Machinery, New York, NY, USA, 235–238. https://doi.org/10.1145/641007.641053

[15] Aurélien Géron. 2019. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Newton, MS, United States.

[16] Hristijan Gjoreski, Ivana Kiprijanovska, Simon Stankoski, Stefan Kalabakov, John Broulidakis, Charles Nduka, and Martin Gjoreski. 2021. Head-Ar: Human Activity Recognition with Head-Mounted IMU Using Weighted Ensemble Learning. In *Activity and Behavior Computing*, Md Atiqur Rahman Ahad, Sozo Inoue, Daniel Roggen, and Kaori Fujinami (Eds.). Springer Singapore, Singapore, 153–167. https://doi.org/10.1007/978-981-15-8944-7_10

[17] Martin Gjoreski, Vito Janko, Gašper Slapničar, Miha Mlakar, Nina Reščič, Jani Bizjak, Vid Drobnič, Matej Marinko, Nejc Mlakar, Mitja Luštrek, et al. 2020. Classical and deep learning methods for recognizing human activities and modes of transportation with smartphone sensors. *Information Fusion* 62 (2020), 47–62.

[18] Nils Y. Hammerla and Thomas Plötz. 2015. Let's (Not) Stick Together: Pairwise Similarity Biases Cross-Validation in Activity Recognition. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (New York, NY, USA, 2015) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 1041–1051. https://doi.org/10.1145/2750858.2807551

[19] Morgan Harvey, Marc Langheinrich, and Geoff Ward. 2016. Remembering through Lifelogging: A Survey of Human Memory Augmentation. *Pervasive and Mobile Computing* 27 (2016), 14–26. https://doi.org/10.1016/j.pmcj.2015.12.002

[20] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Wood. 2006. SenseCam: A Retrospective Memory Aid. In *UbiComp 2006: Ubiquitous Computing*, Paul Dourish and Adrian Friday (Eds.). Vol. 4206. Springer Berlin Heidelberg, Berlin, Heidelberg, 177–193. https://doi.org/10.1007/11853565_11

[21] Md Shafiqul Islam, Tahera Hossain, Md Atiqur Rahman Ahad, and Sozo Inoue. 2021. Exploring Human Activities Using eSense Earable Device. In *Activity and Behavior Computing*, Md Atiqur Rahman Ahad, Sozo Inoue, Daniel Roggen, and Kaori Fujinami (Eds.). Springer Singapore, Singapore, 169–185. https://doi.org/10.1007/978-981-15-8944-7_11

[22] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Allesandro Montanari. 2018. Earables For Personal-Scale Behavior Analytics. *IEEE Pervasive Computing* 17, 3 (2018), 83–89.

[23] Fahim Kawsar, Chulhong Min, Akhil Mathur, Alessandro Montanari, Utku Günay Acer, and Marc Van den Broeck. 2018. ESense: Open Earable Platform for Human Sensing. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems* (New York, NY, USA, 2018) *(SenSys '18)*. Association for Computing Machinery, New York, NY, USA, 371–372. https://doi.org/10.1145/3274783.3275188

[24] Seungchul Lee, Chulhong Min, Alessandro Montanari, Akhil Mathur, Youngjae Chang, Junehwa Song, and Fahim Kawsar. 2019. Automatic Smile and Frown Recognition with Kinetic Earables. In *Proceedings of the 10th Augmented Human International Conference 2019* (New York, NY, USA, 2019) *(AH2019)*. Association for Computing Machinery, New York, NY, USA, Article 25, 4 pages. https://doi.org/10.1145/3311823.3311869

[25] Roya Lotfi, George Tzanetakis, Rasit Eskicioglu, and Pourang Irani. 2020. A Comparison between Audio and IMU Data to Detect Chewing Events Based on an Earable Device. In *Proceedings of the 11th Augmented Human International Conference* (New York, NY, USA, 2020) *(AH '20)*. Association for Computing Machinery, New York, NY, USA, Article 11, 8 pages. https://doi.org/10.1145/3396339.3396362

[26] Chulhong Min, Alessandro Montanari, Akhil Mathur, Seungchul Lee, and Fahim Kawsar. 2018. Cross-Modal Approach for Conversational Well-Being Monitoring with Multi-Sensory Earables. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (Singapore Singapore, 2018-10-08). ACM, Singapore Singapore, 706–709. https://doi.org/10.1145/3267305.3267695

[27] Andreas C. Müller and Sarah Guido. 2016. *Introduction to Machine Learning with Python: A Guide for Data Scientists* (first edition ed.). O'Reilly Media, Inc, Sebastopol, CA, United States.

[28] Jay Prakash, Zhijian Yang, Yu-Lin Wei, and Romit Roy Choudhury. 2019. STEAR: Robust Step Counting from Earables. In *Proceedings of the 1st International Workshop on Earable Computing* (London, United Kingdom, 2019-09-09). ACM, London, United Kingdom, 36–41. https://doi.org/10.1145/3345615.3361133

[29] Tobias Röddiger, Daniel Wolffram, David Laubenstein, Matthias Budde, and Michael Beigl. 2019. Towards Respiration Rate Monitoring Using an In-Ear Headphone Inertial Measurement Unit. In *Proceedings of the 1st International Workshop on Earable Computing* (London, United Kingdom, 2019-09-09). ACM, London, United Kingdom, 48–53. https://doi.org/10.1145/3345615.3361130

[30] Abigail J. Sellen and Steve Whittaker. 2010. Beyond Total Capture: A Constructive Critique of Lifelogging. *Commun. ACM* 53, 5 (May 2010), 70–77. https://doi.org/10.1145/1735223.1735243

[31] Stephanie Tan, David MJ Tax, and Hayley Hung. 2021. Multimodal Joint Head Orientation Estimation in Interacting Groups via Proxemics and Interaction Dynamics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (PACM IMWUT)* 5, 1 (2021), 1–22.

[32] Minda Tessler and Katherine Nelson. 1994. Making memories: The influence of joint encoding on later recall by young children. *Consciousness and cognition* 3, 3-4 (1994), 307–326.

[33] Zi Wang, Sheng Tan, Linghan Zhang, Yili Ren, Zhi Wang, and Jie Yang. 2021. EarDynamic: An Ear Canal Deformation Based Continuous User Authentication Using In-Ear Wearables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–27.

[34] Xuhai Xu, Haitian Shi, Xin Yi, WenJia Liu, Yukang Yan, Yuanchun Shi, Alex Mariakakis, Jennifer Mankoff, and Anind K. Dey. 2020. EarBuddy: Enabling On-Face Interaction via Wireless Earbuds. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA, 2020-04-21). ACM, Honolulu, HI, USA, 1–14. https://doi.org/10.1145/3313831.3376836

Designing Memory Aids for Dementia Patients using Earables

Matija Franklin
University College London
London, UK
matija.franklin@ucl.ac.uk

David Lagnado
University College London
London, UK
d.lagnado@ucl.ac.uk

Chulhong Min
Nokia Bell Labs
Cambridge, UK
chulhong.min@nokia-bell-labs.com

Akhil Mathur
Nokia Bell Labs
Cambridge, UK
akhil.mathur@nokia-bell-labs.com

Fahim Kawsar
Nokia Bell Labs
Cambridge, UK
fahim.kawsar@nokia-bell-labs.com

ABSTRACT

Globally around 50 million people are currently living with dementia, and there are nearly 10 million new cases every year. The decline of memory and, with it, lack of self-confidence and continuous confusion have a devastating effect on people living with this disease. Dementia patients even struggle to accomplish mundane chores and require assistance for daily living and social connectedness. Over the past decade, we have seen remarkable growth in wearable technologies to manage our health and wellbeing and improve our awareness and social connectedness. However, we have to ask why wearables are not addressing this fundamental challenge of memory augmentation that threatens our society? Some limited existing work on cognitive wearables for dementia has focused on using images via camera-based life-logging technology. Instead, in this paper, we argue that *earable* - by virtue of its unique placement, rich sensing modalities, and acoustic feedback capabilities, uncovers new opportunities to augment human cognition to address this pressing need to assist dementia patients. To this end, we delve into fundamental principles of cognitive neuroscience to understand what constitutes memory disorder and its symptoms concerning errors in everyday activities. Building on this, we discuss the benefits of earables (in conjunction with smart objects) in modelling activity and intention of dementia patients and providing contextual memory cues. We put forward a guidance system to assist dementia patients with daily living and social connectedness.

CCS CONCEPTS

• **Computer systems organization → Embedded systems**.

KEYWORDS

earables, dementia, cognitive impairment, memory aids

ACM Reference Format:
Matija Franklin, David Lagnado, Chulhong Min, Akhil Mathur, and Fahim Kawsar. 2021. Designing Memory Aids for Dementia Patients using Earables. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3460418.3479324

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

ACM ISBN 978-1-4503-8461-2/21/09…$15.00
https://doi.org/10.1145/3460418.3479324

1 INTRODUCTION

50 million people worldwide live with dementia, and nearly 10 million new cases are reported every year [1]. The loss of memory and, with it, a sense of identity are the most distressing aspects of this disease. Dementia patients even struggle to accomplish mundane chores and require assistance for daily living and social connectedness [1].

Various memory aids have been proposed to help people with dementia and mild cognitive impairments (MCI). The most common examples of traditional methods are post-it notes, calendars, and diaries [2]. Memory assistive technologies with electronic devices are also widely used, such as note-taking on digital devices and intelligent digital assistants (Siri, Cortana, Google Assistant). Recent wearable technologies enabled context-aware, automated intervention by leveraging a variety of sensors. The wearable remembrance agent [25] is a first-kind-of wearable system for augmented memory. It continuously monitors information retrieved from computers such as note files and emails, and provides just-in-time one-line summaries of information relevant to the user's location and time. Hodges et al. presented SenseCam [11], a wearable camera for a retrospective memory aid, which automatically records surrounding events and helps wearers review the recordings and stimulate their memory. DejaView [4] is a healthcare system that infers a user's surrounding contexts with a combination of sensors, including an accelerometer, a microphone, and a camera, and aids recall of daily activities and plans by unobtrusively cueing the user with relevant information. However, the existing work on cognitive wearables has been still limited to be practically used in daily life. The camera-based solutions for memory support require users to carry a bulky device and to make additional attention to the real-time interaction, e.g., with the smartphone or a computer display.

Here, we ask how wearables can help dementia patients live a better and more independent life. To answer the questions, we focus on the everyday tasks of dementia patients, such as tea-making, getting dressed, and pill-taking, which adults with dementia are known to struggle with. This inability to carry out activities of daily living is associated with a diminished quality of life, poor self-esteem, anxiety, and social isolation [3].

In this paper, we argue that *earable* - by virtue of its unique placement, rich sensing modalities, and acoustic feedback capabilities uncovers new opportunities to augment human cognition to address this pressing need to assist dementia patients. To this end, we delve into fundamental principles of cognitive neuroscience to understand what constitutes this memory disorder and its symptoms concerning errors in everyday activities. We first give a primer on dementia, along with a taxonomy of cognitive issues related to dementia as well as the characteristics of patient errors that result from memory impairments. Then, building on these findings and their implications, we discuss the benefits of earables (in conjunction with smart objects) in modelling activity and intention of dementia patients and providing practical and contextual memory cues. We also put forward a guidance system to assist dementia patients for their daily living, and social connectedness.

2 COGNITION AND DEMENTIA: A PRIMER

We begin by delving into principles of human cognition, cognitive decline, and its relationship with dementia.

Cognitive decline: Age-associated cognitive decline is the inevitable process of normal, non-pathological neurological ageing [16]. From early adulthood, processing ('thinking') speed, working memory, reasoning and executive function all start to decline [5]. The rate of 'normal' brain ageing and the associated cognitive decline depend on many factors, including genetics, general health, lifestyle (e.g., diet), medical disorders, and biological processes such as inflammation. Dementia and mild cognitive impairment (MCI), on the other hand, are relatively rare in that most older people do not develop dementia; with current estimates suggesting that less than one in five people over the age of 80 have dementia [24].

Dementia and cognitive decline: Dementia is a syndrome that is associated with a deterioration of memory and thinking, and an overall decline of cognitive abilities at a greater pace [20]. Symptoms include problems with planning and doing tasks in the right order, memory loss, mood and personality changes, and confusion. Dementia is diagnosed when these symptoms cause problems with activities of daily living (ADLs) to the point that a person cannot live independently. Dementia is not a singular disease but associated with multiple symptoms of memory loss, thinking and communication issues. Alzheimer's disease is the most common type of dementia, making up 60-75% of the total patients [24].

Dementia and MCI are not a part of normal brain ageing, but rather, diagnosable conditions. MCI affects 5-20% of the population aged 65 and over [24]. It disrupts the same cognitive functions affected by 'normal' brain ageing - processing speed, working memory, reasoning and executive function - but to a greater extent. Common functional memory problems reported by people with MCI include forgetting names, numbers and passwords, misplacing things, issues with remembering what was said or decided upon, and keeping track of commitments and intended activities. Thus, MCI does not fully prevent independent living and some cases are treatable. One in 6 cases of MCI progress to dementia within a year.

Dementia and cognitive functions: To understand the cognitive functions affected by dementia, we offer a short overview of cognitive domains and their implications associated with dementia.

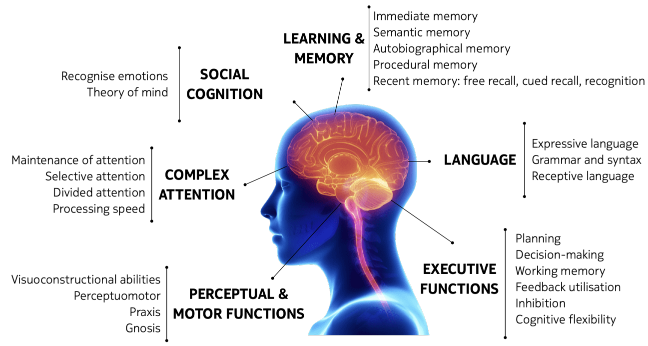


Figure 1: Cognitive domains and corresponding functions associated with dementia and mild cognitive impairment

The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (American Psychiatric Association, 2013) - the taxonomic and diagnostic tool used by the American Psychiatric Association for psychiatric diagnoses - identifies six cognitive domains associated with dementia [12] as illustrated in Figure 1.

- **Complex attention:** This domain covers the maintenance of attention, selective attention, divided attention and processing speed. Disruptions to this ability mean that normal tasks take longer, especially in the presence of competing stimuli.
- **Executive functioning:** This domain reflects the functions concerning planning, decision-making, working memory, feedback utilisation, inhibition and cognitive flexibility. Disruptions to this ability include difficulties with multi-stage tasks, multi-tasking, following directions, organising, planning and keeping up with shifting situations and conversations.
- **Learning and memory:** This domain covers functions of *immediate memory* (e.g., repeating a list of digits), *semantic memory* (e.g., remembering facts), *autobiographical memory* (e.g.., remembering personal events), and *procedural memory* (e.g., recalling skills required to carry out procedures), as well as *recent memory*, which includes free recall (e.g., recalling as many things as possible), *cued recall* (e.g., recalling as many things as possible from a specific category) and recognition. Disruptions to this ability include difficulties recalling recent events, losing track of one's own actions, misplacing objects and repeating oneself.
- **Language:** This domain represents *expressive language* (i.e., fluency in speech), *grammar and syntax*, and *receptive language* (i.e., comprehension). Disruptions to this ability include use of wrong words, grammatical error, word-finding difficulty and difficulties with comprehension of spoken or written language.
- **Perceptual-motor and visuospatial function:** This domain covers *visuoconstructional abilities* (e.g., draw, assemble furniture), *perceptuomotor* (e.g., insert puzzle piece into appropriate slots), *praxis* (e.g., ability to mime gestures) and *gnosis* (e.g., recognises faces and colors). Disruptions in this ability can lead to patients getting lost in familiar places, and finding it difficult to use familiar tools and appliances.
- **Social cognition:** This domain represents the ability to recognise emotions and to have a theory of mind (e.g., considering another person's thoughts and intentions). Disruption to this ability can

lead to a loss of empathy, loss of judgement and inappropriate behaviour.

Dementia patients experience disruptions to these cognitive functions at varying degrees depending on the stage of the disease. However, the common symptoms, irrespective of the decline of specific cognitive functions in dementia patients, are errors in everyday functioning. Naturally, understanding and designing interventions aimed at addressing these symptoms are active research areas. We consider this aspect of understanding and designing interventions for the symptoms is where wearables, and in particular *earable*, can play a critical role. To this end, in the next section, we delve into error patterns typically demonstrated by dementia patients as a structured and systematic understanding of these patterns will provide us with the proper foundation for applying earable-based assistive solutions.

3 UNDERSTANDING ERROR PATTERNS OF DEMENTIA PATIENTS

Areas of cognition that are disrupted by dementia produce well-known patterns of errors. These include calculation, memory of past events, prospective memory (e.g., remembering to attend an upcoming appointment) and the sequencing of complex behaviour [15]. In this section, we reflect on several past studies to systematically identify a set of error patterns associated with dementia patients.

Action coding for error identification: To understand and systematically categorise the different errors produced by dementia patients, researchers have used the action coding system - a method for coding the actions of patients [28]. Here, A1 transcripts provide low-level descriptions of a patient's interaction with the environment that allowed the errors to be identified. There are four different types of actions: move, alter, take, and give. A2 transcripts are procedures within the A1 actions, which can be used to identify errors within the transition between A1 sub-goals. In other words, the flow between different actions can be analysed as well as the degree of overlap between different actions.

A review of research using the action coding system identified eight common dementia patient errors that occur during activities of daily living (ADL), for instance, making a hot drink. The errors - six types of error and two types of incoherent action [28] included:

(1) Place substitution (e.g., putting tea in cereal)
(2) Object substitution (e.g., apple juice added to the cup of tea)
(3) Drinking anticipation (e.g., drinking tea before it is prepared)
(4) Omission errors (e.g., pouring in water from the kettle before it boiled)
(5) Instrumental substitutions (e.g., stirring the tea with a knife)
(6) Faulty execution (e.g., not fully opening a sugar packet)

The incoherent actions were:

(1) Independent acts (e.g., picking up a random item and then putting it down again)
(2) Toying behaviour (e.g., making random gestures with objects with no apparent aim to the action).

A point of note here is that these actions and their transitions can be modelled with activity recognition capabilities afforded by wearables today. We reflect on this in the latter part of this paper.

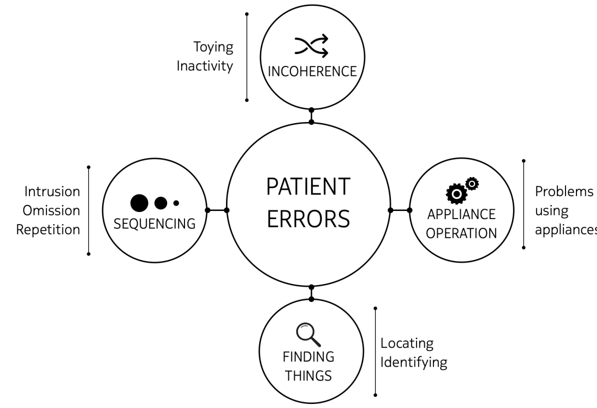


Figure 2: Dementia patient errors in the presence of interventions/active assistance during activities of daily living

Multi-Level action coding for error identification: Researchers have also used the Multi Level Action Test (MLAT) - a standardised test of action disorders - as well as the Naturalistic Action Test (NAT), which is a shortened version of MLAT [26, 27]. In MLAT, the patient being assessed is asked to carry out a task, such as making a slice of toast with jam and butter or wrapping a present. The task is then completed in one of four conditions marked by different levels of difficulty. The first level is *solo basic* where only the required materials are presented. Second is *solo-distractors* where some functionally related items are also presented next to the required materials. Third is *dual-basic* where items for a primary task are presented next to items of another specific task (e.g., making a slice of toast and preparing a cup of tea). Finally, the *dual-sear* condition involves some of the materials which are located in different parts of the room, next to other task-irrelevant items. Research using MLAT has identified eight common dementia patient errors [26].

(1) Omission, which is failing to perform an action
(2) Sequence error, including anticipation-omission (i.e., skipping a necessary step to perform another step), reversal (i.e., reversing the ordering of two steps) and preservation (i.e., performing the action more times than intended)
(3) Object substitution, which is using the wrong object to perform an action
(4) Action addition, where an additional unnecessary action is performed
(5) Gesture substitution, which is performing an action in an uncommon, usually more difficult way
(6) Grasp-spatial misorientation, which involves holding an object in an incorrect way
(7) Spatial misorientation, which is sort of error stemming from misjudgment of size, amount or any similar measure of quantity
(8) Tool omission is using the wrong object to perform an action

Altogether, research finds that omission errors are the most frequent error types, followed by sequence errors [9]. Further, the presence of distractor objects predicatively increases the occurrence of omission and substitution errors.

Error patterns in the presence of interventions: A more ecologically valid study investigated people with dementia performing

ADLs that they themselves identified as important in their own kitchens, including making a cup of tea or coffee [28]. The study tried to assist patients using five levels of prompting: 1) a verbal prompt of the end goal, 2) a verbal prompt of the sub-goal, 3) a verbal prompt of the action, 4) a verbal prompt of action and pointing, 5) performing the action for them. Based on the results, the study concluded four broader areas of error: sequencing, finding things, operation of appliances, incoherence [28] as depicted in Figure 2.

- **Sequencing** errors included intrusion, whereby an inappropriate action is performed from a different activity that prevents the completion of the current activity; omission, whereby a patient misses an action that is required for completing the activity and accomplishing the end goal; repetition, whereby a patient unnecessarily repeats an action that prevents the completion of the activity.
- **Finding things** errors included locating - errors in finding items that are out of view - and identifying - selecting items that are in view.
- **Operation of appliances** errors were problems of using different appliances such as the kettle or toaster.
- **Incoherence** errors included toying - performing random gestures with items with no apparent goal - and inactivity - not performing any action at all.

This systematic analysis of error patterns exhibited by dementia patients provides the foundation for designing intervention solutions with wearables. However, it also expose the requirement for intervention beyond on-body augmentations, i.e., instrumentation of patient environments to offer situated assistance.

Intention-action Gap: The uniformity of results across patients has led researchers to argue that the different types of error all result from a disruption to a cognitive process responsible for goal-directed behaviour [28]. To this end, Norman et al. proposed *the contention scheduling model* - a model of action error - that explains how this disruption would occur [22]. The model proposes that the pathological weakening of top-down activation from a supervisory attentional system means that the contention scheduling system responsible for choosing action schemas is disrupted and does not work the way it should. Bottom-up activation externally from environmental triggers and internally from associated action schemas result in actions that do not follow the intended goal. This produces an intention-action gap in dementia patients.

Implications: We can draw several important implications that emerged from these studies in the design of memory aid applying pervasive technologies to support dementia patients. The contention scheduling model essentially exposes the critical challenge, i.e., to reduce the gap between patient's intention and corresponding action by modelling patient's activity and interaction with the physical world and situational context. This understanding then further be utilised to design interventions applying implicit or explicit memory cues. The error patterns highlight the scope of these challenges quite appropriately.

On **activity modelling**, it is imperative to understand: 1) patients' motion-induced physical activity and gesture, 2) the state and identity of the physical object that a patient interacts with to accomplish the intended task, 3) the exact interaction dynamics concerning the physical object in context.

On **intention modelling**, it is critical to understand the patient's overall objective to derive a plan applying causal reasoning grounded on predictive modelling of a patient's past actions for the same purpose. This aspect uncovers interesting modelling challenges and demands thoughtful mitigation strategy both for directive and corrective guidance.

On **designing memory cues**, we see opportunities in two different dimensions. First, it is imperative to create implicit memory cues, e.g., voice prompts, just-in-time visuals, to direct and correct patients' actions. Second, we also see opportunities to augment physical objects with awareness technologies (sensors and actuators) to augment their functional capabilities to participate in a patient activity in a proactive way.

In the next section, we discuss how we can apply these implications to design memory aids for dementia patients.

4 EARABLE AS MEMORY AID

In the previous sections, we offered a concise overview of cognition decline, its relationship with dementia, and its implications on dementia patients concerning errors in ADLs. We also identified three primary challenges in designing memory aids to support dementia patients in mitigating these errors. In this section, we want to posit that *earables*, together with smart objects, provide the proper foundation for designing assistive guidance for dementia patients. Grounded on *contention scheduling model*, we learned that reducing the gap between intended objective and corresponding actions is one of the critical facets to assist dementia patients. Taking a deeper view of this facet, we have identified three key dimensions: activity modelling, intention modelling, and effective memory cues that can collectively mitigate erroneous actions. Earables today come with rich sensors, including inertial measurement unit (IMU), microphones, Bluetooth Low Energy (BLE), and in some instances, optical sensor (PPG), core body temperature sensor, and electrodermal activity (EDA) sensor. These sensors and their placement in the ear collectively offer us unique opportunities to observe and understand internal (biomarkers) and external contexts around a human body and offer us privacy-preserving, intimate, and subtle feedback capabilities through synthesised speech, music, and acoustic cues. These capabilities are critical to all three dimensions we listed before.

Activity modeling: We can recognise upper body movements, such as head and neck activities [6, 21], facial activities and expression [18], and whole-body movements, i.e., walking, standing, falling, etc. [21]. These motion primitives are vital artefacts to understand a patient's motion-induced physical activities and gestures. Note that modelling all these context primitives, especially around the head, is not possible with other wearables. The acoustic channel of an earable enables us to understand environment ambience and audio events [19, 23], thereby modelling a patient's proxemic, social context as well patient's interaction with physical objects [10]. Combining these primitives and their thoughtful synthesis is key in modelling patient activities and creating digital memories through encoding for future recall.

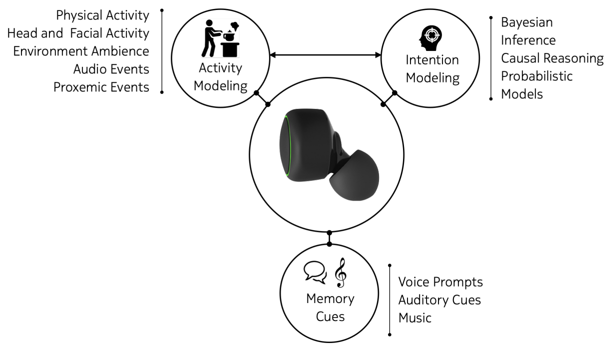


Figure 3: Unique opportunities of earables with context primitives from activity and intention modelling, and acoustic feedback to design effective memory cues

Intention modelling : These context primitives also hold the key to causal reasoning to understand the patient's intention. Of course, a patient can use explicit instructions, such as *"I want to make a cup of tea*, or *"I want to have my medication"* to express her intention. However, we argue these context primitives are the causal link to decipher a patient's intended action. For instance, head orientation and gaze indicate an activity location where a patient might be interested; picking a particular object with a distinct soundscape or moving to a specific direction might tell a patient's intention to a broader activity. Causal models that could draw inferences about the expected activity in a particular location can eliminate or reduce confusion that a dementia patient experiences. We advocate further research on Bayesian techniques, causal inferences, and probabilistic models grounded on these context primitives to model patient's intentions. In addition, we also see opportunities to model the errors that a patient encounters to predict potential divergence from an intended activity. Using different context primitives, an earable can accurately represent a patient's situational context and erroneous actions. We can later exploit these actions to identify an early indication of a potential mistake.

Memory Cues: Finally, earable offers a unique opportunity to provide memory cues using its acoustic channel. Literature on human memory demonstrated that auditory stimuli remain in our sensory registry for at least 4 seconds (compared to 1 second of visual stimuli). This aspect is very critical for dementia patients due to their declined cognition, as we have explained before. Given the privacy-preserving and intimate placement and delivery mechanism of earables, synthesised voice prompts, auditory cues, and music can be designed as memory cues to guide dementia patients. This feature is a big differentiation attribute of earable compared to other on-body wearables with memory aids for dementia patients.

5 SMART OBJECT AS MEMORY CUES

Earables can offer implicit memory cues; however, as discussed earlier, modelling the activity of a dementia patient also demands an accurate understanding of object interaction. Over the past decade, we have seen remarkable progress in smart object research in which everyday objects are instrumented with awareness technologies, i.e., sensors and actuators, to offer value-added functionalities beyond their primary established purposes [7, 8, 13, 17]. We consider these objects in concert with earables to provide the best foundation for designing memory aids for dementia patients. We envision physical objects that can understand their state of use and can proactively activate to offer auditory or visual cues (e.g., playing a tone, voice prompts, or illumination) to guide the activities of dementia patients and offer the activity traces that we can leverage to build a better causal reasoning model to understand a patient's intention.

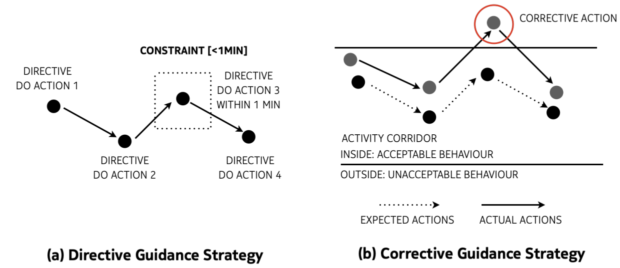


Figure 4: Two different strategies to guide dementia patients

6 AMBIENT GUIDANCE SYSTEM

Building on the activity models, intention models, and memory cues we discussed in the earlier section, here we present a blueprint of an ambient guidance system combining earables and smart objects as memory aids for dementia patients. Our system borrows principle from Situated Flows [14], first reported by Kawsar et al. to design an activity-aware situated guidance system for workers in a structured workplace. A situated flow (flow, for short) is a high-level declarative abstraction for modelling real-life processes and human activities. It consists of a set of actions glued together by a plan (a set of transitions), which defines how actions should be performed to achieve some goal under a set of constraints. In the context of this work, a situated flow essentially describes the prescribed steps of an activity that a patient is interested in accomplishing. These flows can be predefined or derived from the patients' past actions while completing a task.

Our ambient guidance system consists of two components: guidance strategies for deciding which information should be accessible and when, where, and how it should be presented in the patient's immediate environment using earables and smart objects. Situated flows represent context-specific prescriptions for how activities and tasks are supposed to be done or how a dementia patient operates an appliance. Memory cues with earables and smart objects make it possible to expose activity and task information to a patient. However, to effectively guide a patient, it is not enough to present a patient with every single step. Practical guidance requires a guidance strategy that defines:

- Which tasks and activities are exposed to a patient.
- When and where guidance information is presented.
- How to present guidance information with memory cues.
- How to cope with situations in which a patient does not follow the guidance.

In order to cope with such disparate requirements, we propose two levels of generic guidance strategies.

Directive guidance: Directive guidance (Figure 4(a)) is a strategy that presents a patient with just-in-time notifications (directives) of the following activities to be done. To be precise, directives are generated and presented to a patient before an activity has to be performed. For example, in a medication context, before and during taking medicines, it is beneficial to provide an updated (if any) instruction to the patient.

Corrective guidance: Corrective guidance (Figure 4(b)) is a strategy that assumes that a patient has a sufficient understanding of what she has to do and that she does not require constant reminders. Instead, this strategy only presents a patient with guidance information when the system detects significant deviations from the plan. This aspect is visualized in Figure 4(b): an activity corridor defines how much an actual activity may deviate from the one prescribed by the activity plan. If an action falls outside the activity corridor, the system issues corrective feedback to inform patients of the deviation and motivate the patient to follow the plan as described. For example, in a tea-making scenario, if a patient accidentally picks a salt canister instead of a sugar canister, the guidance system kicks in with a reminder. The corrective plans can be dynamically generated from the activity model and current activity state.

One may argue that, instead of earables, the proposed guidance system can be built with cameras installed in care homes and video analytics solution. It might be able to provide higher accuracy of activity and intention modelling, but cannot provide instant, contextual memory cues as effectively as earables. Also, they are inherently limited to be used in everyday situations due to privacy invasion, narrow spatial coverage, expensive hardware cost.

7 CONCLUDING REMARKS

Dementia is a threat to our aging population. Cognitive deficits and a loss of self-identity of dementia patients fundamentally challenge our society to react and rethink. Wearables, and in particular earables, offer a unique opportunity to contribute to the ongoing effort in addressing this societal challenge. The present article aims to provide theoretical and methodological insights that provide a solid foundation for wearable technology. We first gave a primer on dementia, along with a taxonomy of cognitive issues related to dementia as well as the characteristic patient errors that result from memory impairments. Then, building on these findings and their implications, we discussed the benefits of earable (in conjunction with smart objects) in modeling activity and intention of dementia patients and providing practical and contextual memory cues. We also put forward a guidance system to assist dementia patients. Finally, we would like to experimentally validate this earable-based guidance system in multiple ecological valid studies to uncover its efficacy in our work's future avenue.

REFERENCES

[1] [n.d.]. Fact sheets: Dementia, World Health Organization. https://www.who.int/news-room/fact-sheets/detail/dementia. Accessed: June 22, 2021.

[2] Robin N Brewer, Meredith Ringel Morris, and Siân E Lindley. 2017. How to remember what to remember: exploring possibilities for digital reminder systems. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1, 3 (2017), 1–20.

[3] Alistair Burns and Peter Rabins. 2000. Carer burden in dementia. *International Journal of Geriatric Psychiatry* 15, S1 (2000), S9–S13.

[4] Dirk De Jager, Alex L Wood, Geoff V Merrett, Bashir M Al-Hashimi, Kieron O'Hara, Nigel R Shadbolt, and Wendy Hall. 2011. A low-power, distributed, pervasive healthcare system for supporting memory. In *Proceedings of the First ACM MobiHoc Workshop on Pervasive Wireless Healthcare*. 1–7.

[5] Ian J Deary, Janie Corley, Alan J Gow, Sarah E Harris, Lorna M Houlihan, Riccardo E Marioni, Lars Penke, Snorri B Rafnsson, and John M Starr. 2009. Age-associated cognitive decline. *British medical bulletin* 92, 1 (2009), 135–152.

[6] Andrea Ferlini, Alessandro Montanari, Cecilia Mascolo, and Robert Harle. 2019. Head motion tracking through in-ear wearables. In *Proceedings of the 1st International Workshop on Earable Computing*. 8–13.

[7] Kaori Fujinami, Fahim Kawsar, and Tatsuo Nakajima. 2005. Awaremirror: A personalized display using a mirror. In *International Conference on Pervasive Computing*. Springer, 315–332.

[8] H-W Gellersen, Albrecht Schmidt, and Michael Beigl. 2000. Adding some smartness to devices and everyday things. In *Proceedings Third IEEE Workshop on Mobile Computing Systems and Applications*. IEEE, 3–10.

[9] Tania Giovannetti, David J Libon, Laurel J Buxbaum, and Myrna F Schwartz. 2002. Naturalistic action impairments in dementia. *Neuropsychologia* 40, 8 (2002), 1220–1232.

[10] Taesik Gong, Hyunsung Cho, Bowon Lee, and Sung-Ju Lee. 2019. Knocker: Vibroacoustic-based object recognition with smartphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–21.

[11] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Wood. 2006. SenseCam: A retrospective memory aid. In *International Conference on Ubiquitous Computing*. Springer, 177–193.

[12] Julie Hugo and Mary Ganguli. 2014. Dementia and cognitive impairment: epidemiology, diagnosis, and treatment. *Clinics in geriatric medicine* 30, 3 (2014), 421–442.

[13] Fahim Kawsar, Kaori Fujinami, and Tatsuo Nakajima. 2005. Augmenting everyday life with sentient artefacts. In *Proceedings of the 2005 joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies*. 141–146.

[14] Fahim Kawsar, Gerd Kortuem, and Bashar Altakrouri. 2010. Supporting interaction with the internet of things across objects, time and space. In *2010 internet of things (IOT)*. IEEE, 1–8.

[15] Alexandra C King and Corinna Dwan. 2019. Electronic memory aids for people with dementia experiencing prospective memory loss: a review of empirical studies. *Dementia* 18, 6 (2019), 1994–2007.

[16] Arpita Konar, Padmanabh Singh, and Mahendra K Thakur. 2016. Age-associated cognitive decline: insights into molecular switches and recovery avenues. *Aging and disease* 7, 2 (2016), 121.

[17] Gerd Kortuem, Fahim Kawsar, Vasughi Sundramoorthy, and Daniel Fitton. 2009. Smart objects as building blocks for the internet of things. *IEEE Internet Computing* 14, 1 (2009), 44–51.

[18] Seungchul Lee, Chulhong Min, Alessandro Montanari, Akhil Mathur, Youngjae Chang, Junehwa Song, and Fahim Kawsar. 2019. Automatic Smile and Frown Recognition with Kinetic Earables. In *Proceedings of the 10th Augmented Human International Conference 2019*. 1–4.

[19] Dawei Liang and Edison Thomaz. 2019. Audio-based activities of daily living (adl) recognition with large-scale acoustic embeddings from online videos. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–18.

[20] Gill Livingston, Jonathan Huntley, Andrew Sommerlad, David Ames, Clive Ballard, Sube Banerjee, Carol Brayne, Alistair Burns, Jiska Cohen-Mansfield, Claudia Cooper, et al. 2020. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet* 396, 10248 (2020), 413–446.

[21] Chulhong Min, Akhil Mathur, and Fahim Kawsar. 2018. Exploring audio and kinetic sensing on earable devices. In *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications*. 5–10.

[22] Donald A Norman and Tim Shallice. 1986. Attention to action. In *Consciousness and self-regulation*. Springer, 1–18.

[23] Chunjong Park, Chulhong Min, Sourav Bhattacharya, and Fahim Kawsar. 2020. Augmenting Conversational Agents with Ambient Acoustic Contexts. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–9.

[24] Sujata Ray and Susan Davidson. 2014. Dementia and cognitive decline A review of the evidence. *Age UK* 27, 10–12.

[25] Bradley J Rhodes. 1997. The wearable remembrance agent: A system for augmented memory. *Personal Technologies* 1, 4 (1997), 218–224.

[26] Myrna F Schwartz, Michael W Montgomery, Laurel J Buxbaum, Sonia S Lee, Tania G Carew, H Coslett, Mary Ferraro, Eileen Fitzpatrick-DeSalme, Tessa Hart, and Nathaniel Mayer. 1998. Naturalistic action impairment in closed head injury. *Neuropsychology* 12, 1 (1998), 13.

[27] Myrna F Schwartz, Mary Segal, Tracy Veramonti, Mary Ferraro, and Laurel J Buxbaum. 2002. The Naturalistic Action Test: A standardised assessment for everyday action impairment. *Neuropsychological rehabilitation* 12, 4 (2002), 311–339.

[28] Joseph P Wherton and Andrew F Monk. 2010. Problems people with dementia have with kitchen tasks: The challenge for pervasive computing. *Interacting with Computers* 22, 4 (2010), 253–266.

Towards Automatic Recognition of Perceived Level of Understanding on Online Lectures using Earables

Dongwoo Kim[*]
NAVER
Seongnam, Republic of Korea
woo.kim@navercorp.com

Chulhong Min[†]
Nokia Bell Labs
Cambridge, UK
chulhong.min@nokia-bell-labs.com

Seungwoo Kang
KOREATECH
Cheonan, Republic of Korea
swkang@koreatech.ac.kr

ABSTRACT

The COVID-19 pandemic has seriously impacted education and forced the whole education system to shift to online learning. Such a transition has been readily made by virtue of today's Internet technology and infrastructure, but online learning also has limitations compared to traditional face-to-face lectures. One of the biggest hurdles is that it is challenging for teachers to instantly keep track of students' learning status. In this paper, we envision earables as an opportunity to automatically estimate learner's understanding of learning material for effective learning and teaching, e.g., to pinpoint the part for which learners need to put more effort to understand. To this end, we conduct a small-scale exploratory study with 8 participants for 24 lectures in total and investigate learner's behavioral characteristics that indicate the level of understanding. We demonstrate that those behaviors can be captured from a motion signal on earables. We discuss challenges that need to be further addressed to realize our vision.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; • **Applied computing** → *E-learning*.

KEYWORDS

Online Learning, Understanding Level, Earable, Automatic Recognition

ACM Reference Format:

Dongwoo Kim, Chulhong Min, and Seungwoo Kang. 2021. Towards Automatic Recognition of Perceived Level of Understanding on Online Lectures using Earables. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3460418.3479323

[*]This work was done while the first author was with KOREATECH.
[†]Corresponding author.

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

ACM ISBN 978-1-4503-8461-2/21/09…$15.00
https://doi.org/10.1145/3460418.3479323

1 INTRODUCTION

The COVID-19 pandemic has impacted every aspect of our lives. Among others, education has been seriously affected. Schools in a lot of countries had to close or reduce their face-to-face classes. According to some reports [1, 2], more than 1 billion students worldwide could not use their classrooms at the peak of the crisis. Alternatively, schools are providing access to education using online learning technology and numerous students are currently educated remotely all over the world. The pandemic would accelerate the educational innovation far beyond the advance of online learning witnessed over the last decade, e.g., Coursera and Udacity.

While the transition to online learning has been readily made by virtue of today's Internet technology and infrastructure, online learning also has limitations compared to traditional face-to-face lectures. One of the biggest hurdles is that teachers could not instantly keep track of students' learning status. In pre-recorded video lectures such as Coursera, teachers can neither observe how students engage in lectures, e.g., nonverbal and behavioral cues, nor interact with them. In live lectures using online conferencing tools such as Zoom, such observation and interaction can be possible if students have a camera and a microphone, but it imposes significant burdens on teachers, especially when there are a number of students in a lecture. These limitations hinder teachers from adapting their lecture materials or teaching methods when necessary. One typical method is to give a quiz after/during a lecture, but it is also burdensome to teachers due to the time and effort required.

In this paper, we envision *earables* (also known as smart earbuds) as an opportunity to automatically estimate learner's understanding status of learning materials. Such functionality would enable effective learning and teaching even in online lectures, e.g., to pinpoint the part for which learners need to put more effort to understand. From an explorative study, we uncover that a learner's postures and head motions can be a clue to represent their understanding of lectures. We present the capability and feasibility of identifying learners' understanding levels based on such behavioral patterns. Then, we discuss opportunities and challenges for realizing *the automatic estimation of understanding* in the wild.

2 RELATED WORK

Previous studies have demonstrated the capability of detecting online learners' diverse status (e.g., inattention, engagement, frustration, mind wandering) based on their behavioral patterns [3, 7, 13–15]. For example, Mota et al. presented a technique to recognize naturally occurring postures of learners and detect their interest level based on pressure sensors mounted on a chair [13]. Pham et al. proposed a multimodal approach to infer learners' affective and cognitive states such as boredom, confusion, and frustration

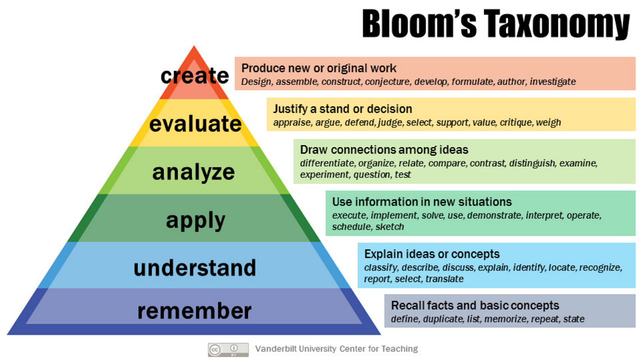


Figure 1: Bloom's Taxonomy (the image by the Vanderbilt University Center for Teaching / CC-BY [1])

by analyzing facial expressions and PPG signals captured from the front and back cameras of a smartphone [14]. Robal et al. presented an IntelliEye system that tracks face and mouth use of online learners to detect in-attention state [15]. It also provides alerts to them when they are in-attentive to return learners' attention to a lecture video. Grafsgaard et al. proposed to predict the engagement and frustration of students during computer-mediated tutoring based on fined-grained facial movements [7]. Bosch et al. proposed an automatic mind wandering detector by analyzing a set of facial features from face videos of students [3].

These works commonly imply that online learners' behaviors during the lecture carry meaningful information about their learning status. However, they mostly rely on physiological sensors and/or computer vision, thereby significantly limiting their practicality in real-life situations, due to the need for using additional, dedicated devices and privacy concerns. We investigate how learners' behaviors, more specifically postures and motion gestures, can be interpreted to represent their understanding while taking an online lecture. For example, a student might often tilt her head or look at a monitor vacantly when she does not understand what she listens to. Similarly, some students might nod their head when they well follow what is taught. Our initial idea and preliminary study was presented in [9].

3 APPROACH TO UNDERSTANDING ESTIMATION

3.1 Modelling Understanding

Measuring how well students learn and understand lectures is essential for teachers to provide effective teaching. Bloom's Taxonomy is a widely-adopted educational framework, which was developed to assist teachers to plan classes and design valid assessment strategies [4, 10]. The revised version of Bloom's Taxonomy introduces six levels of cognitive learning; remembering, understanding, applying, analyzing, evaluating, and creating [10]. As shown in Figure 1, each level represents different cognitive skills and learning behaviors from the most basic to the more complex levels. For example, *remembering* is related to retrieving, recalling, and recognizing factual information and relevant knowledge, and *understanding* is related to interpreting, summarizing, and explaining main ideas and concepts of learning material. Moving up the levels, they refer to higher cognitive thoughts and skills. In this study, we adopt the first two levels of Bloom's Taxonomy, remembering and understanding$^+$, to model online student's understanding* of learning material [2].

Table 1: Questionnaire for the level of understanding.

1	I could tell the important keywords/concept of the lecture
2	I could briefly explain the important keywords/concept of the lecture
3	I could tell what I newly learned
4	I could explain the summary of the lecture content
5	I could explain the lecture content so that others can understand it

3.2 Quantifying Understanding Level

To quantify the level of learner's understanding, we design a questionnaire by adopting Bloom's Taxonomy. We note that teachers are often encouraged to use different types of questions in class and on assignments and tests based on Bloom's Taxonomy to stimulate and assess students' cognitive thinking. Example questions that can be used are as follows.

- How would you define ... ?
- What was the main idea ... ?
- Can you write a brief outline ... ?
- Can you provide an example of ... ?

Inspired by these, we design the questionnaire with five statements as shown in Table 1. We construct the first two statements in the table for the concept of remembering and the rest three for the concept of understanding$^+$. For each statement, respondents are asked to rate how much they agree with the statement using a 5-point Likert scale (1 to 5) (5 - "strongly agree", 4 – "agree", 3 – "neutral", 2 – "disagree", 1 – "strongly disagree").

We use Likert scale answers that can be easily collected and quantified regardless of lecture types and contents. Asking detailed answers specific to the lecture content might be better to assess respondent's understanding level more accurately, but it imposes much burden on respondents to answer and assessors. Investigating the effect of different types of questions will be our future work.

As a granularity of understanding level estimation, we target *a lecture slide* as a unit, i.e., estimating the student's understanding level for each lecture slide. A slide often conveys a single topic and content, and thus it is naturally expected by teachers to be mapped the level of learner's understanding. Other lecture units, e.g., an explanation of a specific term in the slide or a whole lecture, can also be considered for different purposes. We leave it future work.

3.3 Why Motion Sensing on Earables?

A key decision to be made for the design of a sensing solution is to determine devices and sensors to be used. Many existing methods often rely on computer vision using a user-facing camera to detect learner's behaviors, thereby significantly limiting their applicability in real-life situations and raising privacy concerns. Unlike them, we focus on motion signals (accelerometer and gyroscope)

[1]https://www.flickr.com/photos/vandycft/29428436431

[2]Understanding$^+$ refers to the specific level of Bloom's Taxonomy. We use understanding* as the term representing both *remembering* and *understanding* of the taxonomy. In the rest of the paper, understanding refers to understanding*, unless otherwise noted.

on earable devices. Our choice offers several benefits. First, according to our study in the following section, postures and gestures relevant to understanding are mostly made around the head and upper body, which could be captured by earable devices. Second, processing motion signal is computationally efficient and privacy preserving compared to other methods, especially computer vision. Last but not least, earables are widely used when students take online lectures and inertial measurement unit (IMU) for motion signal is already employed on most smart earbuds. Thus, an earbud-integrated motion-sensing solution would be easily adopted without requiring additional devices.

4 BEHAVIORAL CUES FOR UNDERSTANDING ESTIMATION

4.1 DATA COLLECTION

For an in-depth study, we exploit a dataset including 8 participants and 24 online lectures in total. The participants (4 males and 4 females) were recruited from a university campus, and they were graduate and undergraduate students in Computer Science and Engineering. Their ages were between 23-26 (mean: 24.13, SD: 3.75). Each participant was compensated with a gift card equivalent to USD 18. The study was approved by the Institutional Review Board of KOREATECH (No. 20022502).

Table 2 shows the online lectures we used for the study. We choose four lectures on the course of Artificial Intelligence Basic, provided by K-MOOC, a Korean MOOC established in 2015. All the participants have a general interest in the topic of AI, but did not take lectures with the same content as the lectures in our study.

Each participant was invited to the lab for data collection. We explained the purpose and procedure of the study and obtained informed consent. Based on each participant's prior knowledge and level on AI and ML, we chose three different lectures that cover a range of difficulty levels. During the lecture, the participants were asked to wear the eSense earbuds [8, 12] for sensor data collection. They were also provided with printed lecture materials and a pen to make them feel that they take lectures as usual.

During the lecture, we collected three types of data from the participants: (1) 3-axis accelerometer and 3-axis gyroscope data sampled at 32 Hz from eSense to analyze their behaviors while taking online lectures, (2) a video stream using a participant-facing webcam as ground truth of their behaviors, and (3) questionnaire (Table 1) answers on every slide as ground truth of their understanding. The participants completed the questionnaire after finishing each lecture. Between lectures, they took a break of 5 minutes. From the responses to the questionnaire, we obtain a final understanding score between 5 and 25 for a lecture slide, by summing all the scores from five answers, with 25 indicating the highest possible understanding score.

4.2 Natural Behaviors during Online Lectures

We extract a set of behaviors that learners naturally make during online lectures. For the analysis, the researchers transcribed and coded the recorded videos with observable, repetitive behaviors that the participants performed while they were watching the lectures.

Figure 2 shows the list of behaviors we observed in the videos (See the left side of the figure); we excluded infrequently observed behaviors. We group them into three categories; *posture*, *body/head motion*, and *facial motion*. As expected, learners' macroscopic, whole-body movements were quite limited during online lectures, mostly sitting and watching the video. However, interestingly, a variety of their microscopic movements were observed, especially around the head and upper body. From this behavioral characteristic, we believe that our choice of earbuds as a sensing device has a great potential to capture learners' behaviors on online lectures.

4.3 Understanding-relevant Behaviors

As a next step, we identify key behaviors that can be used as clues to estimate understanding level. To understand the impact of our observed behaviors, we assess the statistical association between a) the statistic of observed behaviors and b) the reported scores of understanding in the questionnaire. For the statistic of the observed behaviors, we annotate the start and end time of behavior events and compute the statistical features on every slide. For the behaviors that last for a certain time duration, we measure the duration of every segments of behaviors and normalize it by dividing by the slide duration. Note that we additionally compute the total, average and standard deviation for the posture behaviors that last longer than 1 second. For the behaviors involving a brief motion, we count the number of events in every slide and also normalize it with the slide duration. We use Spearman's rank correlation for the correlation analysis, which is used to assess the relationship between two variables measured on at least an ordinal scale [11].

Figure 2 shows the correlation coefficients of the behavior statistics and the reported understanding scores (See the right side). The results show two important findings. First, there are several behavioral patterns that imply a moderate relationship to the level of understanding, e.g., over 0.3 or under -0.3. This implies the potential of estimating a learner's *understanding* from the observation of behavioral patterns. Interestingly, after the survey, we heard from our participants some comments that support our analysis. For example, P7 reported that he frequently looked at the lecture material when he could not understand well. Accordingly, he frequently lowered and raised the head. Such patterns are observed through the coefficient analysis, e.g., the negative correlation of the count of *briefly gazing at a monitor*, *lowering the head*, and *raising the head*. P6 mentioned that he often did neck rolls and changed his sitting posture when he could not understand well. These behavioral patterns are revealed in the negative correlation of the count of *moving the neck* and *briefly moving the body*.

Second, the relationship between behavioral patterns and the level of understanding differs depending on the individual, showing the need for personalized estimation. For example, five behaviors, i.e., *keeping looking down a desk*, *keeping gazing at a monitor*, *briefly gazing at a monitor*, *lowering the head*, and *nodding*, show correlation coefficients larger than 0.3 or smaller than −0.3 for P7, but they do not for P4. P4 has only one behavior, *moving the neck*, that shows a correlation coefficient larger than 0.3.

4.4 Feasibility of Understanding Estimation

To study the feasibility of understanding level estimation, we build regression models to predict the understanding score using the aforementioned behavioral features. Since the impact of behavioral

Table 2: Online lectures used in data collection.

Course	Topics	Duration	# of slides	# of participants
Artificial Intelligence Basic	Introduction to Reinforcement Learning	34 min.	13	8
	Markov Process	22 min.	10	8
	Markov Decision Process	39 min.	13	7
	Heuristic Search	22 min.	17	1

Category	Behaviors	Features (every slide)	P1	P2	P3	P4	P5	P6	P7	P8
Posture	Keeping looking down a desk (>= 1 sec)	Total duration	-0.06	-0.01	-0.08	0.04	0.22	0.37*	0.01	0.15
		Average duration	-0.11	-0.03	-0.18	0.00	0.47*	0.47*	0.33*	0.25
		Std of duration	-0.23	-0.16	0.02	-0.19	0.13	0.39*	0.21	0.12
	Briefly looking down a desk briefly (<1 sec)	Count of events	0.13	0.14	-0.25	0.15	-0.15	-0.15	-0.24	-0.01
	Keeping gazing at a monitor (>= 1 sec)	Total duration	0.02	0.03	0.14	-0.01	-0.20	-0.35*	0.11	-0.14
		Average duration	-0.01	0.09	0.15	-0.14	0.13	-0.05	0.38*	0.16
		Std of duration	-0.08	-0.02	0.36*	-0.09	-0.11	-0.24	0.15	-0.17
	Briefly gazing at a monitor (< 1 sec)	Count of events	0.00	-0.07	-0.31	-0.02	-0.04	-0.02	-0.36*	-0.26
Body/head motion	Lowering the head	Count of events	0.16	-0.19	-0.22	0.05	0.18	0.16	-0.45*	0.02
	Raising the head	Count of events	0.37*	0.10	-0.23	0.07	0.15	0.14	-0.23	0.05
	Leaning the body forward	Count of events	0.21	0.12	-0.09	0.14	-0.13	0.01	-0.01	-0.19
	Leaning the body backward	Count of events	0.03	0.30	-0.21	0.14	-0.10	-0.29	0.02	-0.24
	Leaning the body to the right	Count of events	-0.16	-0.09	-0.14	-0.13	0.00	-0.06	0.17	-0.24
	Leaning the body to the left	Count of events	0.23	0.09	-0.33*	-0.05	0.00	-0.08	0.17	-0.14
	Nodding	Total duration	0.20	0.18	0.05	-0.13	-0.02	-0.14	-0.33*	-0.18
	Moving the neck	Total duration	-0.15	0.02	0.20	0.42*	-0.29	-0.25	-0.15	-0.01
	Shaking legs	Total duration	0.00	0.01	0.00	0.00	0.27	-0.12	0.20	0.00
	Touching the face	Count of events	-0.08	0.02	0.05	-0.10	-0.02	-0.31	0.23	-0.19
	Briefly moving the body	Total duration	0.56*	-0.00	0.03	-0.04	-0.26	-0.25	0.21	-0.39*
Facial motion	Grimacing	Count of events	0.02	0.00	0.00	0.00	0.00	-0.21	0.00	-0.06
	Yawning	Count of events	-0.03	0.13	-0.02	0.21	-0.06	-0.39*	0.16	0.00
	Sighing	Count of events	0.19	-0.28	-0.00	0.00	0.00	-0.00	0.00	0.00
	Laughing	Count of events	0.00	0.00	-0.16	0.00	0.00	-0.09	0.00	0.00
	Dozing	Total duration	0.00	0.00	-0.30	0.00	0.06	0.00	0.00	0.00

Figure 2: Behavior list and correlation coefficients of behavior features and understanding score (* indicates p-value < 0.05)

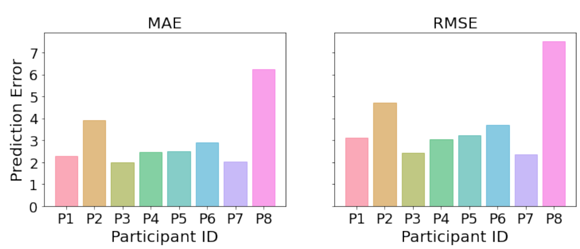


Figure 3: Prediction error for overall understanding scores

patterns is different depending on the individual, we train a separate regression model for each participant. We adopt linear regression. To train the model, we use the features with top-3 correlation coefficient values for each participant. We apply leave-one-slide-out cross validation.

Figure 3 shows the prediction errors with two metrics, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Our estimation model achieves 3.04 of the average MAE and 3.77 of the average RMSE across the participants; the target value is the total understanding level score and its range is from 5 to 25. For comparison, we use a baseline that takes an average understanding score as a prediction. The average MAE and RMSE of the baseline are 3.35 and 3.96, respectively. Our model shows the better estimation results for both metrics. We additionally compare the results of SVM regression with different types of kernels, i.e., linear and RBF, using top-3 features. They show slightly larger errors for both metrics, around 3.2 and 4 of MAE and RMSE, respectively.

We can also observe the variation of prediction errors depending on the participant. For example, P1, P3 and P7 show around 2 of MAE while P2 and P8 show 3.9 and 6.2 of MAE, respectively. We find that those who show relatively high errors have the smaller number of features with larger coefficient values. Also, their range of understanding scores is relatively larger across the lecture slides than others. We discuss this issue in the following section.

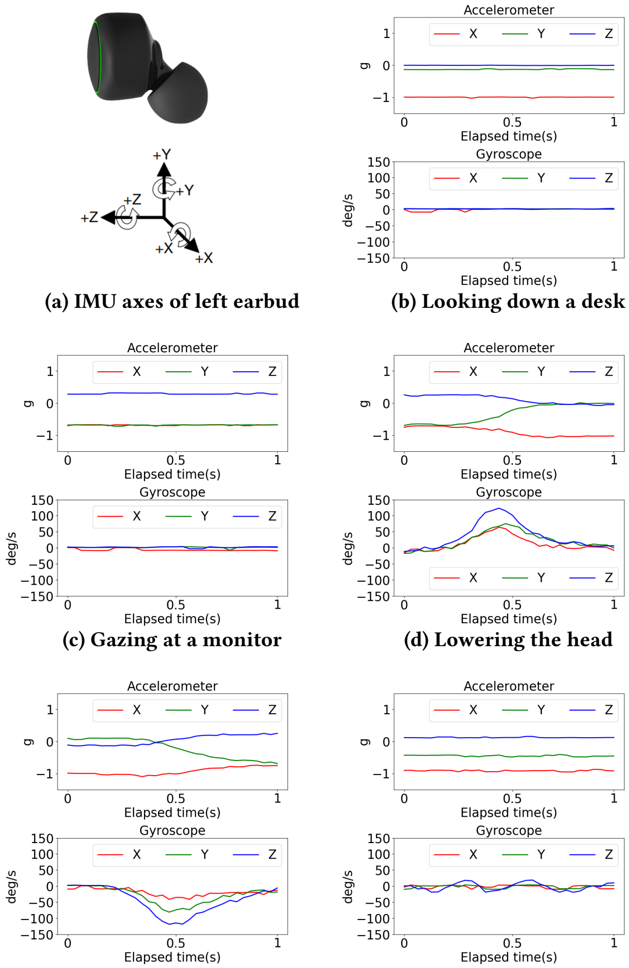
5 OPPORTUNITIES AND CHALLENGES

Our explorative study shows that the automatic estimation of learner's understanding is promising from identifying understanding-relevant behaviors and mapping them to the understanding score. In this section, we discuss opportunities and challenges for realizing the automatic estimation of understanding in the wild.

5.1 Behavior Detection using Earables

A key for the automatic estimation of understanding is to detect understanding-relevant behaviors at runtime. Here, we explore detection techniques for these behaviors using earable devices and their performance.

As an initial attempt, we choose five behaviors as primitive contexts for automatic estimation of learner's understanding level: two postures (*looking down a desk* and *gazing at a monitor*) and three motion gestures (*lowering the head*, *raising the head*, and *nodding*). These behaviors show a meaningful relationship with the understanding score, i.e., correlation coefficients larger than 0.3 or smaller than -0.3 in many participants and also observed relatively more frequently than other behaviors throughout the lectures. Note that, by detecting two postures, we can derive all of the posture category features in Figure 2.



(a) IMU axes of left earbud

(b) Looking down a desk

(c) Gazing at a monitor

(d) Lowering the head

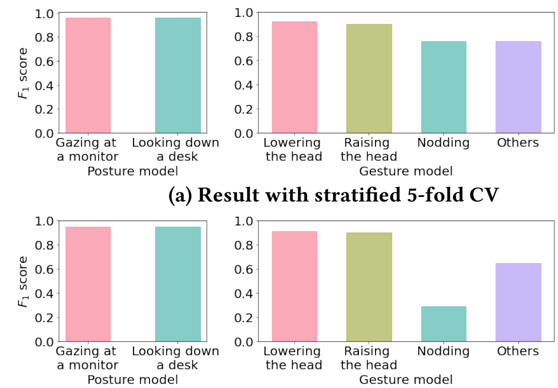
(e) Raising the head

(f) Nodding

Figure 4: Motion signal of different postures and gestures

Detection technique: Postures and gestures have distinctive characteristics of motion signal patterns. Figure 4 shows the accelerometer and gyroscope data for some examples. While signals show little change over time while a user is taking a posture (e.g., Figure 4b and 4c), the fluctuation of signals can be easily observed in motion gestures (e.g., Figure 4d, 4e, and 4f). Inspired by such a finding, we devise a two-stage sensing pipeline to detect understanding-relevant behaviors. In the first stage, the pipeline quantifies the degree of movement and identifies whether a given signal segment is from a posture or a gesture. Then, in the second stage, it employs two machine learning models, one for posture detection and the other for gesture detection, and selectively uses them based on the output of the first stage. For each task, posture or gesture detection, a single model is built for all the users.

Movement detection: We take one-second gyroscope samples as input and quantifies the movement by calculating the signal variation. More specifically, we compute the magnitude value of every 3-axis gyroscope sample and calculate the variance of 32 values. The higher variance represents the higher degree of movement. Then, we distinguish between a posture and a gesture using a threshold that we empirically set using our dataset.



(a) Result with stratified 5-fold CV

(b) Result with leave-one-subject-out CV

Figure 5: Recognition performance

Machine learning models: One of the challenges in distinguishing between *gazing at monitor* and *looking down a desk* is that learners mostly remain stationary when they make these behaviors. Accordingly, we can easily expect that traditional activity pipelines for smartphones [16] would not work well because they are mostly designed to utilize the magnitude stream as input to address the arbitrary position of smartphones. On the contrary, the relative direction of the earbuds to a person's head is mostly fixed. Thus, we can leverage the absolute orientation of a device. We empirically found that X-axis and Z-axis show a strong discrimination power to identify understanding-relevant behaviors.

We segment accelerometer and gyroscope data streams into one second-frame. Then, we extract time and frequency-domain features [5] from X and Z streams separately without taking the magnitude and gather the features for the classification. We use PCA to reduce the dimensionality of the features and Support Vector Machine (SVM) as a classifier; we fine-tuned hyper-parameters using our dataset. Two machine learning models have the same architecture, but different target labels. The posture model is to separate looking down a desk and gazing at a monitor, and the gesture model is to separate lowering the head, raising the head, nodding, and others. We do not include the *others* label in the posture model because other postures are hardly observed in our data collection setup. However, we believe our model can be easily extended to cover other postures if needed.

Results and implications: We investigate the recognition performance of our detection technique. We conduct a stratified 5-fold cross-validation and report the F_1 score as a performance metric. The experimental results show that our machine learning models detect understanding-relevant behaviors accurately. Figure 5a show the F_1 score of posture and gesture models, respectively. The posture model shows 0.96 of F_1 score for both postures, looking down a desk and gazing at a monitor. It validates our choice of using axis-specific streams as input, instead of the magnitude. The gesture model also shows reasonable performance. The F_1 scores of two gestures, lowering and raising the head, are 0.92 and 0.90, respectively. However, the F_1 scores of nodding and others are relatively lower, 0.76 for both. This was mainly because the gesture model sometimes confuses the events of nodding and others.

5.2 Variation of Behavioral Patterns

One of the challenges for deploying the automatic estimation of understanding to end-users is the variety of behavioral patterns. First, our participants have a different set of behaviors that show a meaningful relationship with their understanding. For example, while the average duration of the *looking down a desk* posture shows a high correlation for P5, P6, and P7, but has little relationship for P1, P2, and P4.

Second, the signal patterns of behaviors are also different depending on the participant. To investigate how our posture and gesture models work on a new user, we measure their performance with a leave-one-subject-out validation (Figure 5b). The results show that the posture model still achieves high accuracy, i.e., 0.95 of F_1 score. The gesture model also shows the reasonable performance for lowering and raising the head, i.e., 0.91 and 0.90, respectively, which are expected to have little variation across the participants. However, the performance of nodding and others largely decreases, i.e., 0.29 and 0.65 of F_1 score. Observing the collected data, we could see that the participants often did nodding and other behaviors differently, e.g., in terms of direction, count, and strength.

These two findings imply the need for personalized models, i.e., detection model for behavior recognition and regression model for estimating the understanding score. We believe we can develop personalized models using a small amount of the end-user's data with online learning techniques. We leave it as future work.

5.3 Incorporating Additional Sensors

This study currently explores the feasibility of using behavioral features from earables' IMU to estimate students' understanding. It would be possible to incorporate additional sensors for more accurate and robust estimation. Some previous works utilize physiological sensors such as PPG and EDA to detect engagement or attention state of students [6]. While they are different from our target, understanding level, these might be related to each other, considering that engagement in lectures might positively affect the understanding of the lectures. We will further investigate the potential of adopting sensor fusion techniques to use physiological features from PPG and EDA sensors. A further in-depth study will also be necessary to analyze the relationship between the understanding level and the engagement or attention state.

5.4 Application Landscape

We envision that the automatic estimation of learner's understanding will provide significant benefits in online lectures.

Our proposed technique could be used to assist teachers by providing student's understanding status. In many online lectures, especially the pre-recorded video lectures, it is almost infeasible for teachers to have detailed real-time feedback from students, thereby making it difficult to modify and enhance their lectures; even possible, the feedback is often at a high, coarse-grained level. We envision that our solution can monitor student's understanding status, gather this information, and automatically spot parts in online lectures which student perceive difficult to understand.

Our solution could also help students by calling their attention when they do not understand the lecture content well. When our solution detects such moments, it could send a notification message to the students, e.g., "how about watching this part again if you do not understand well?". It might also be possible to provide the students with a summary of difficult parts after the lecture in order to help the student reflect on the lecture.

6 CONCLUSION

This work is our initial step towards a vision to utilize earables as an opportunity to automatically estimate learner's understanding on online lectures. In this explorative study, we observe some behaviors that could be used as clues to estimate the understanding level, and investigate the feasibility of understanding estimation using these behaviors. To realize the automatic estimation of understanding in the wild, we have a range of challenges to address, which will be future avenues of our work.

ACKNOWLEDGMENTS

This research was supported in part by the Ministry of Science and ICT, Korea, under the ITRC support program (IITP-2020-0-01778) supervised by the IITP.

REFERENCES

[1] [n.d.]. *The COVID-19 pandemic has changed education forever. This is how.* Retrieved June 14, 2021 from https://www.weforum.org/agenda/2020/04/coronavirus-education-global-covid19-online-digital-learning/
[2] [n.d.]. *Education during COVID-19; moving towards e-learning.* Retrieved June 14, 2021 from https://www.europeandataportal.eu/en/impact-studies/covid-19/education-during-covid-19-moving-towards-e-learning
[3] Nigel Bosch and Sidney D'Mello. 2019. Automatic detection of mind wandering from video in the lab and in the classroom. *IEEE Transactions on Affective Computing* (2019).
[4] Patricia A Eber and Trent S Parker. 2007. Assessing Student Learning: Applying Bloom's Taxonomy. *Human Service Education* 27, 1 (2007).
[5] Davide Figo, Pedro C Diniz, Diogo R Ferreira, and Joao MP Cardoso. 2010. Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing* 14, 7 (2010), 645–662.
[6] Nan Gao, Wei Shao, Mohammad Saiedur Rahaman, and Flora D Salim. 2020. n-Gage: Predicting in-class Emotional, Behavioural and Cognitive Engagement in the Wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–26.
[7] Joseph Grafsgaard, Joseph B Wiggins, Kristy Elizabeth Boyer, Eric N Wiebe, and James Lester. 2013. Automatically recognizing facial expression: Predicting engagement and frustration. In *Educational Data Mining 2013*.
[8] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Alesandro Montanari. 2018. Earables for personal-scale behavior analytics. *IEEE Pervasive Computing* 17, 3 (2018), 83–89.
[9] Dongwoo Kim, Chulhong Min, and Seungwoo Kang. 2020. Towards recognizing perceived level of understanding for online lectures using earables. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 739–740.
[10] David R Krathwohl. 2002. A revision of Bloom's taxonomy: An overview. *Theory into practice* 41, 4 (2002), 212–218.
[11] Ann Lehman. 2005. *JMP for basic univariate and multivariate statistics: a step-by-step guide.* SAS Institute.
[12] Chulhong Min, Akhil Mathur, and Fahim Kawsar. 2018. Exploring audio and kinetic sensing on earable devices. In *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications*. ACM, 5–10.
[13] Selene Mota and Rosalind W Picard. 2003. Automated posture analysis for detecting learner's interest level. In *2003 Conference on Computer Vision and Pattern Recognition Workshop*, Vol. 5. IEEE, 49–49.
[14] Phuong Pham and Jingtao Wang. 2017. AttentiveLearner 2: a multimodal approach for improving MOOC learning on mobile devices. In *International Conference on Artificial Intelligence in Education*. Springer, 561–564.
[15] Tarmo Robal, Yue Zhao, Christoph Lofi, and Claudia Hauff. 2018. IntelliEye: Enhancing MOOC Learners' Video Watching Experience through Real-Time Attention Tracking. In *Proceedings of the 29th on Hypertext and Social Media*. ACM, 106–114.
[16] Aras Yurtman. 2019. *Activity recognition invariant to position and orientation of wearable motion sensor units.* Ph.D. Dissertation. bilkent university.

AirCase: Earable Charging Case with Air Quality Monitoring and Soundscape Sonification

Haibin Zhao
hzhao@teco.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Tobias Röddiger
roeddiger@teco.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Michael Beigl
beigl@teco.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

ABSTRACT

Bad air quality and insufficient ventilation can have severe impacts on personal health. We present AirCase, a smart earable charging case that measures CO_2, volatile organic compounds, humidity, air pressure, temperature, and light intensity. The case powers both the air quality system and the earables. We also propose a model-driven air quality soundscape sonification strategy based on the audio capabilities of the earables. AirCase detects conditions unsuitable for measuring air quality (e.g., in pocket) in an office environment at 98.2 % accuracy with a simple classifier based on a single feature. We identified light intensity as the primary indicator to recognize occlusion. In contrast, the speed of the micro ventilator used to increase airflow inside the case did not offer any predictive value. In the future, we hope to see more researchers explore the hidden potential of the new platform.

CCS CONCEPTS

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

KEYWORDS

earables; hearables; earphones; charging case; air quality monitoring; pollution

ACM Reference Format:
Haibin Zhao, Tobias Röddiger, and Michael Beigl. 2021. AirCase: Earable Charging Case with Air Quality Monitoring and Soundscape Sonification. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3460418.3479329

1 INTRODUCTION

Earables have been heavily explored as a novel, ear-worn platform with personal tracking capabilities. Some examples are monitoring health-related parameters such as respiration [18] or brain activity [9]; or activity recognition to detect daytime activities [14] or eating episodes [4]. The platform also serves as the foundation of new interaction paradigms [25, 28]. In research, the focus, therefore, is on the device worn on the ear itself. State-of-the-art market-available earables are usually wireless and sold together with a charging case that contains much larger batteries compared to the earbuds[1]. Today, these charging cases are solely intended to provide power to the earables when the user is not wearing them, so they are recharged for the next use. However, as the user always carries the case with them, it creates an opportunity to implement additional sensors for mobile sensing. As the charging case may be placed, e.g., next to a person on the table, it seems particularly suitable for mobile environmental sensing scenarios.

To initially make the earable community aware of the hidden potential of the earable charging case and to spark novel ideas, we built a first prototype that follows the new paradigm. The charging case measures air quality and shares it back to the user by a soundscape sonification strategy. We conducted a data collection study in three different conditions (non-occluded + bright, non-occluded + dark, and occluded + dark) to gather 150 minutes of air quality data (30 x 5 minutes). Based on the collected data, we trained a simple classifier and performed 5-fold-cross-validation yielding 98.2 % overall accuracy.

In sum, our three main contributions are:

- an off-the-shelf charging cases with custom electronics to measure CO_2, volatile organic compounds (VOC), humidity, air pressure, temperature, and the light intensity, and which includes a micro ventilator for improved airflow in AirCase
- a model-based soundscape sonification strategy of air quality on earables
- an occlusion detection classifier to avoid measuring air quality when the device is, e.g., in the pocket

2 BACKGROUND AND RELATED WORK

Air quality and human life are closely coupled. For example, high humidity and temperature affect the level of human comfort [23] and high concentrations of VOC and CO_2 in the air even harm human health [10, 15]. Consequently, insufficient room ventilation rates in offices have severe adversarial effects such as higher rates of respiratory infections and overall more short-term sick leaves [22]. Therefore, personal mobile air quality measurement devices are expected to help the individual live a healthier life.

There have already been some mobile air quality monitoring devices in the field of ubiquitous computing. For example, *WearAir* [12] installed a VOC sensor and light-emitting diodes (LEDs) to a T-shirt. The LEDs indicate the air quality around the user measured by the VOC sensor, which informs people about the air quality

[1] https://www.apple.com/airpods-pro/

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

ACM ISBN 978-1-4503-8461-2/21/09…$15.00
https://doi.org/10.1145/3460418.3479329

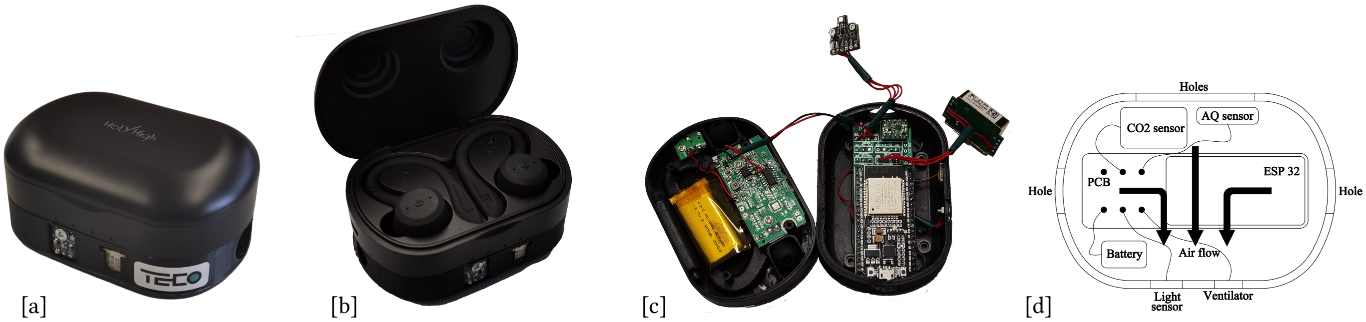


Figure 1: Different views of AirCase. [a] front side of the device showing the light sensor (left) and micro-ventilator (right); [b] open lid showing the earables placed inside the case for charging; [c] opened case showing the charging circuit (left) and custom PCB connected to sensors (right); [d] schematic drawing of the sensing tier.

straightforward. Jiang et al. [11] proposed a device equipped with CO_2, humidity, temperature, and light sensors, which can be attached on the backpack or clothes of the user. It performs indoor localization based on WiFi signals and can project the air quality onto the correct room. In addition to carrying sensor nodes directly on the human, there is also research focusing on vehicle-mounted air quality sensor nodes [3, 24]. The abundant space on vehicles generally leaves room for a larger number of more accurate sensors to be integrated than with wearables.

Related work in earable computing embedded environmental sensors directly inside the ear-worn device, but not for the primary purpose of air quality monitoring. E.g., to prevent heat strokes, temperature and humidity sensors inside earables can measure the body temperature, and evaporation of the wearer [7, 13]. Though air pressure sensors have been embedded inside earables, their sole purpose was to serve as an underlying sensing principle for interaction such as face gestures [1] or input by the muscles inside the ear directly [17].

To the best of our knowledge, no air quality sensors were embedded into earables or their related devices. Therefore, we proposed AirCase (Figure 1), an unobtrusive, augmented earphone charging case with air quality sensing for daily life.

3 AIRCASE

AirCase builds upon an off-the-shelf earbud charging case that we modified by adding custom electronics. Air quality sonification is presented to the user by the respective earables of the case. Occlusion detection relies on a simple machine learning classifier. Figure 2 gives an overview of the system components: in terms of information flow, multiple sensors are mounted to the MCU that runs an occlusion classifier. Measurements are then emitted via BLE to the mobile phone for recording and further processing, including sonification. From the aspect of energy flow, the battery of the charging case serves not only the earbuds but also the MCU and its sensors as well as the ventilation.

3.1 Hardware

Figure 1 shows the structure of AirCase. The foundation is the HolyHigh® earphones charging case (regular wireless Bluetooth earbuds with microphones and a push-button on each side, see Figure 1 [b]). A light sensor (SHARP® GA1A1S202WP) is installed on the front side of AirCase, aiming to detect the light intensity for the occlusion recognition. In addition, a ventilator (SUNON® UB393-500) is equipped next to the light sensor, which serves not only the occlusion recognition (measuring its rotation speed) but also the ventilation of AirCase, as the sensors located inside it must reflect the air quality of the surroundings. To improve ventilation, we have also designed holes on the back and sides of AirCase. To measure air quality, we placed an air quality sensor (BOSCH® BME680) to sense humidity, air pressure, temperature, and volatile organic compounds (VOC). As the BME680 only returns VOC as resistance, we obtain the indexed air quality (IAQ) using the Bosch Sensortec Environmental Cluster (BSEC) library based on VOC and takes temperature and humidity into account. This allows us to assign VOC air quality based on the reference IAQ values provided by Bosch. In addition, we added a CO_2 sensor (WINSEN® MH-Z19B). We utilized the battery of the charging case as the power supply of the aforementioned sensors and the micro-controller unit (MCU), an ESP32. The firmware is written in Arduino. For easier assembly, we designed a custom ESP32-based printed circuit board (PCB) that sits tightly in the case and simplifies cable management (see Figure 1 [c], right). The rough overhead of AirCase is listed in Table 1. We remove the screws from the case and 3D-print a new bottom lid which is slightly thicker to account for the space required for the sensors. To emit, record, and analyze the acquired signals, we developed an Android application to communicate with

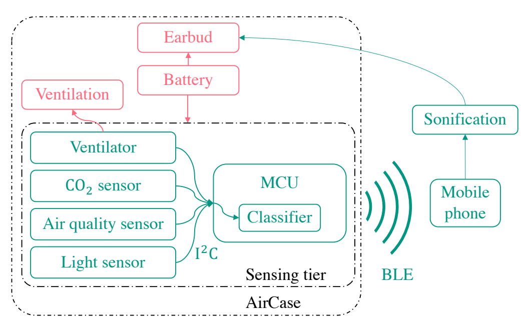


Figure 2: Frame of AirCase system. Red parts present the energy flow and green parts indicate the information flow.

the ESP32 via Bluetooth Low Energy (BLE). The measurements are sampled and sent every five seconds (0.2 Hz). Figure 1 [d] shows the schematic drawing of AirCase.

In the extreme case, i.e. with all sensors active and constantly sending data via Bluetooth, the AirCase can operate continuously for 5 hours. Thanks to the occlusion detection (see subsection 3.3), the AirCase will run for more than 5 hours, depending on the state the AirCase is in, as the MCU stops collecting and sending data to a certain extent in occlusion.

Table 1: System overhead.

Hardware	Cost
Earbud + charging case	40 €
Light sensor	3 €
Ventilator	15 €
Air quality sensor	15 €
CO_2 sensor	25 €
Customized PCB	1 €
Summation	99 €

3.2 Soundscape Sonification Strategy

Earables offer a design space for rich and immersive soundscape experiences. Sonification, on the other hand, describes the transformation of data to sound. In AirCase, we designed a two-tiered soundscape sonification strategy: a background tier, indicating temperature as well as humidity referring to human comfort and an alerting tier to inform the user about potentially harmful air quality with respect to the level of VOC (measured as IAQ) and CO_2.

Background Tier. To combine humidity and temperature into a single measure, we defined the *modified temperature* T_m as proposed by Steadman [21]:

$$T_m = 1.07T + 0.2e - 2.7\,,$$

where T indicates the measured temperature and

$$e = 6.105\frac{H}{100}\exp\left(\frac{17.27T}{237.27+T}\right)$$

with H denoting the humidity. Based on a study by Tanabe and Kimura [23], we define the *neutral temperature* T_n with respect to the modified temperature by averaging the neutral temperature for different groups of users as:

$$T_n = 25.7^\circ C\,.$$

We use the absolute difference between T_m and T_n to control the volume of the music playing on the earables. The greater the absolute difference is, the higher the volume is.

Alerting Tier. The alerting tier informs the user about deviations in air quality that have a significant impact on human health, i.e., VOC and CO_2. The alarm starts when CO_2 value exceeds 1000 ppm [10] or VOC values above 100 IAQ (described as "little bad air quality" by the manufacturer of the VOC sensor [6]). The higher the level of CO_2 and VOC are, the higher the volume of the alerts will be.

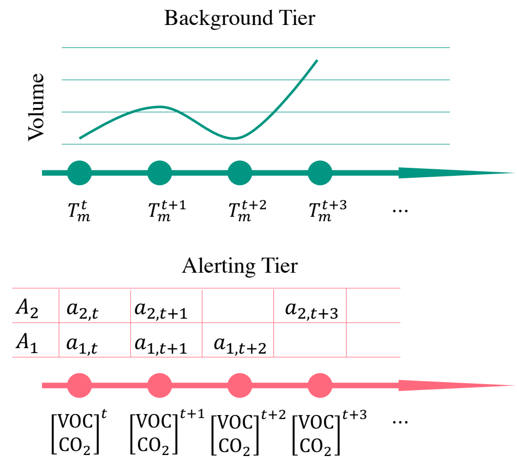


Figure 3: Sonification model of AirCase. Background tier gives the user feedback about comfort and the alerting tier warns the user about potentially harmful air quality.

Sonification Model. Wolf et al. [26] proposed a model for data-driven sonification using soundscapes that we apply to our use case. Figure 3 shows our approach. For the background tier, the volume of background music changes according to changes of the modified temperature at each time step. For the alerting tier, VOC and CO_2 activate the respective alerts (i.e., A_1 and A_2). $a_{i,t}$ denote the alarm in group A_i at time step t. In our implementation, we choose t as one minute and apply a heavy breathing and coughing sound to warn about CO2 and VOC levels, respectively. The user can choose background music according to their preferences as we manipulate the system settings directly.

3.3 Occlusion Recognition

When AirCase is occluded (e.g., while in the bag or pocket of the wearer), the readings of the sensors are not informative as they do not reflect the air quality of the surroundings. Keeping the measurements active under such conditions is also a waste of energy. The main differences between occlusion and an open environment are light and ventilation, as the occlusion blocks the light and impedes the airflow. Thus, we were interested to realize occlusion recognition based on the measurements of light intensity and rotation speed of the ventilator, which we initially explored in our paper.

Dataset. We collected 150 minutes overall in 15 occluded and 15 non-occluded conditions at our lab (5 minutes recording each). Occlusions include placing AirCase in pockets (N=5), bags (N=5), and cabinets (N=5) of five different users. We sampled data at ten different rooms at different times (1 / 3 at night and 2 / 3 during the day) to change light settings. As we want to evaluate occlusion more robustly on a per-minute basis, we applied a one-minute sliding window on every 5-minute recording (step size 5 seconds).

Feature Extraction. Before classifying the samples, we performed automated feature extraction with *tsfresh*. We apply it to the light intensity and ventilator speed time series (each series lasts 1 minute) and filter out the most relevant features associated with labels based on the false discovery rate [5]. This resulted in no relevant features based on the speed of the ventilator. This was also confirmed by training classifiers on the features of the ventilator that did no perform better than random (see Table 2). Looking at the remaining

Table 2: Occlusion detection using ventilator speed.

Classifier	Precision	Recall	F1	Accuracy
SVM	71.7%	99.6%	83.4%	81.2%
Neural Network	69.0%	89.0%	78.0%	75.0%
Random Forest	69.0%	7.4%	13.4%	51.5%

features quickly revealed that the absolute light intensity is the sole predictor of occlusion in our dataset. The feature that represented this best was the absolute sum of light intensity of the sample. Figure 4 [a] shows the logarithmic of s_l denoting the sum of light intensity in sliding windows of all recordings, while [b] shows the logarithmic of s_v, the sum values of ventilator speed in sliding windows. It can be well observed that the light is distinctive to predict the occluded and non-occluded conditions, whereas the ventilator speed does not change consistently across them.

Threshold Classifier. To avoid having to select the optimal threshold by hand, we fit a support vector machine classifier (SVM) using *sklearn*. We perform 5-fold-cross-validation and avoid that samples from the same recording are in the training and validation set at the same time using three different classifiers i.e. SVM, neural network, and random forest. After training and validation, we acquire the result described in Table 3.

Results. From Table 3 we find that a SVM achieves an acceptable classification accuracy in our office environment for occlusion detection using the light sensor measurements solely. Therefore, we can turn off the sensors to save energy when the device is occluded. As the ventilator's speed does not improve occlusion recognition, we can also shut it down.

Table 3: Performance of the AirCase occlusion detection.

Classifier	Precision	Recall	F1	Accuracy
SVM	96.8%	100%	98.4%	98.2%
Neural Network	91.0%	100%	95.3%	95.0%
Random Forest	98.5%	69.2%	81.3%	84.4%

4 DISCUSSION

We have demonstrated how an earable charging case can be equipped with air quality sensing, introduced a soundscape sonification strategy, and preliminary showed the possibility to detect occlusions.

4.1 Limitations

Currently, we only evaluated the occlusion detection of AirCase on a limited set of data samples acquired from the same building. It remains to be investigated how well AirCase works under real-life settings. For example, additional sensors such as a light spectrum sensor may be required. Possibly, the user could define this threshold for occlusion detection based on personal preference, or AirCase could activate depending on the current activity of the user (e.g., detecting where the device is carried [2]).

Our sonification strategy remains to be evaluated with the users.

Also, we would like to emphasize that AirCase is currently much larger than necessary. The BME680 and GA1A1S202WP sensors both only have a footprint of $\leq 3 \times 3 \times 1\ mm^3$ (l × w × h). Therefore, they could be well integrated directly onto the charging case's printed circuit board.

4.2 Air Quality Feedback

At the moment, information about air quality is presented as raw values inside the app or based on our sonification strategy that remains to be evaluated with the user. Another feedback mechanism could present air quality through light directly on the earable that makes the surrounding people in public spaces aware of the information obtained by the case [8]. Of course, the case could also show the air quality directly.

4.3 Smart Charging Cases

Beyond the smart earable charging case with air quality monitoring presented in our paper, we share three other possible research and application paths.

Health Tracking. Compared to a phone, an earable charging case may be kept inside the pocket during usage. Therefore, it could serve as a better sensing platform for continuous health-related parameter tracking. For example, it could predict the energy expenditure and count the steps of the user, as the hip location proved to be most reliable in past work [19]. Also, the surface area of the charging case presents opportunities for sensors. A handheld Electrocardiography measuring device could place its conductive metal electrodes on the outside of the case (e.g., similar to *AliveCor Kardia Mobile*).

Subtle Gestures. Subtle interaction that requires low effort and can be hidden from others continues to receive attention in HCI research and was recently systematically investigated [16]. We imagine that a smart charging case carried in the pocket could be an enabler of such hidden interactions. For example, Saponas et al. [20] and Wu et al. [27] presented pocket-based input based on textile sensors. Capacitive sensors embedded on the outside of a charging case could allow similar interactions.

Item Tracking. Wu et al. [27] introduced a method to recognize objects placed inside the pocket of the user (e.g., earbuds charging case). Similarly, the charging case itself could be equipped with sensors to make it aware of other items inside the pocket. The case could then remind the user about forgotten objects.

5 CONCLUSION

Earable charging cases offer space and large amounts of power for new types of use cases. We explored the potential of the new platform as an unobtrusive mobile air quality sensor station for daily usage. AirCase builds upon a simple threshold-based approach by measuring light intensity to detect occlusions of the case. The speed of the micro-ventilator as an indirect measure of airflow did not have any predictive value. Additionally, AirCase uses the

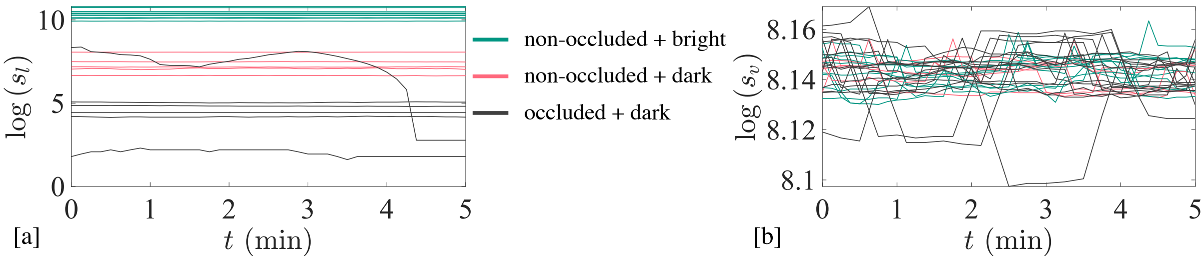


Figure 4: Logarithmic sum of values in sliding windows. [a] logarithmic sum values of light intensity, [b] logarithmic sum values of ventilator speed.

earables' audio capabilities to inform the wearer about air quality based on a model-driven soundscape sonification strategy.

Future Work. We are interested in looking at other strategies for occlusion detection. Moreover, we are looking forward to other work that makes use of the earable charging case as a sensory platform.

REFERENCES

[1] Toshiyuki Ando, Yuki Kubo, Buntarou Shizuki, and Shin Takahashi. 2017. Canalsense: Face-related movement recognition system based on sensing air pressure in ear canals. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 679–689.

[2] Stephen A Antos, Mark V Albert, and Konrad P Kording. 2014. Hand, belt, pocket or bag: Practical activity tracking with mobile phones. *Journal of neuroscience methods* 231 (2014), 22–30.

[3] Paul M Aoki, RJ Honicky, Alan Mainwaring, Chris Myers, Eric Paulos, Sushmita Subramanian, and Allison Woodruff. 2009. A vehicle for research: using street sweepers to explore the landscape of environmental community action. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 375–384.

[4] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. 2017. EarBit: using wearable sensors to detect eating episodes in unconstrained environments. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1, 3 (2017), 1–20.

[5] Yoav Benjamini and Daniel Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics* (2001), 1165–1188.

[6] Bosch. 2017.

[7] Jorge S Chaglla E, Numan Celik, Wamadeva Balachandran, et al. 2018. Measurement of core body temperature using graphene-inked infrared thermopile sensor. *Sensors* 18, 10 (2018), 3315.

[8] Inkyung Choi and Dain Kim. 2020. Toning: New Experience of Sharing Music Preference with Interactive Earphone in Public Space. In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 533–538.

[9] Valentin Goverdovsky, Wilhelm Von Rosenberg, Takashi Nakamura, David Looney, David J Sharp, Christos Papavassiliou, Mary J Morrell, and Danilo P Mandic. 2017. Hearables: Multimodal physiological in-ear sensing. *Scientific reports* 7, 1 (2017), 1–10.

[10] Tyler A Jacobson, Jasdeep S Kler, Michael T Hernke, Rudolf K Braun, Keith C Meyer, and William E Funk. 2019. Direct human health risks of increased atmospheric carbon dioxide. *Nature Sustainability* 2, 8 (2019), 691–701.

[11] Yifei Jiang, Kun Li, Lei Tian, Ricardo Piedrahita, Xiang Yun, Omkar Mansata, Qin Lv, Robert P Dick, Michael Hannigan, and Li Shang. 2011. MAQS: a personalized mobile sensing system for indoor air quality monitoring. In *Proceedings of the 13th international conference on Ubiquitous computing*. 271–280.

[12] Sunyoung Kim, Eric Paulos, and Mark D Gross. 2010. WearAir: expressive t-shirts for air quality sensing. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*. 295–296.

[13] Keiji Matsumoto, Yuksel Temiz, Hamidreza Taghavi, Elrick L Cornelius, Hiroyuki Mori, and Bruno Michel. 2019. An earbud-type wearable (A hearable) with vital parameter sensors for early detection and prevention of heat-stroke. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 7049–7055.

[14] Chulhong Min, Akhil Mathur, and Fahim Kawsar. 2018. Exploring audio and kinetic sensing on earable devices. In *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications*. 5–10.

[15] Lars Mølhave. 1991. Volatile organic compounds, indoor air quality and health. *Indoor Air* 1, 4 (1991), 357–376.

[16] Henning Pohl, Andreea Muresan, and Kasper Hornbæk. 2019. Charting subtle interaction in the hci literature. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–15.

[17] Tobias Röddiger, Christopher Clarke, Daniel Wolffram, Matthias Budde, and Michael Beigl. 2021. EarRumble: Discreet Hands-and Eyes-Free Input by Voluntary Tensor Tympani Muscle Contraction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[18] Tobias Röddiger, Daniel Wolffram, David Laubenstein, Matthias Budde, and Michael Beigl. 2019. Towards Respiration Rate Monitoring Using an In-Ear Headphone Inertial Measurement Unit. In *Proceedings of the 1st International Workshop on Earable Computing*. 48–53.

[19] Ramyar Saeedi, Navid Amini, and Hassan Ghasemzadeh. 2014. Patient-centric on-body sensor localization in smart health systems. In *2014 48th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2081–2085.

[20] T Scott Saponas, Chris Harrison, and Hrvoje Benko. 2011. PocketTouch: through-fabric capacitive touch input. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. 303–308.

[21] Robert G Steadman. 1984. A universal scale of apparent temperature. *Journal of Applied Meteorology and Climatology* 23, 12 (1984), 1674–1687.

[22] Jan Sundell, Hal Levin, William W Nazaroff, William S Cain, William J Fisk, David T Grimsrud, F Gyntelberg, Y Li, AK Persily, AC Pickering, et al. 2011. Ventilation rates and health: multidisciplinary review of the scientific literature. *Indoor air* 21, 3 (2011), 191–204.

[23] Shinichi Tanabe and Kenichi Kimura. 1994. *Effects of air temperature, humidity, and air movement on thermal comfort under hot and humid conditions.* Technical Report. American Society of Heating, Refrigerating and Air-Conditioning Engineers ....

[24] Péter Völgyesi, András Nádas, Xenofon Koutsoukos, and Ákos Lédeczi. 2008. Air quality monitoring with sensormap. In *2008 International Conference on Information Processing in Sensor Networks (ipsn 2008)*. IEEE, 529–530.

[25] Yu-Te Wang, Masaki Nakanishi, Simon Lind Kappel, Preben Kidmose, Danilo P Mandic, Yijun Wang, Chung-Kuan Cheng, and Tzyy-Ping Jung. 2015. Developing an online steady-state visual evoked potential-based brain-computer interface system using EarEEG. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2271–2274.

[26] KatieAnna E Wolf, Genna Gliner, and Rebecca Fiebrink. 2015. A model for data-driven sonification using soundscapes. In *Proceedings of the 20th International Conference on Intelligent User Interfaces Companion*. 97–100.

[27] Te-Yen Wu, Zheer Xu, Xing-Dong Yang, Steve Hodges, and Teddy Seyed. 2021. Project Tasca: Enabling Touch and Contextual Interactions with a Pocket-based Textile Sensor. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[28] Xuhai Xu, Haitian Shi, Xin Yi, Wenjia Liu, Yukang Yan, Yuanchun Shi, Alex Mariakakis, Jennifer Mankoff, and Anind K Dey. 2020. EarBuddy: Enabling On-Face Interaction via Wireless Earbuds. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

Earables for Detection of Bruxism: a Feasibility Study

Erika Bondareva[*]
University of Cambridge
United Kingdom
eb729@cam.ac.uk

Elín Rós Hauksdóttir[*]
University of Cambridge
United Kingdom
erh72@cam.ac.uk

Cecilia Mascolo
University of Cambridge
United Kingdom
cm542@cam.ac.uk

ABSTRACT

Bruxism is a disorder characterised by teeth grinding and clenching, and many bruxism sufferers are not aware of this disorder until their dental health professional notices permanent teeth wear. Stress and anxiety are often listed among contributing factors impacting bruxism exacerbation, which may explain why the COVID-19 pandemic gave rise to a bruxism epidemic. It is essential to develop tools allowing for the early diagnosis of bruxism in an unobtrusive manner. This work explores the feasibility of detecting bruxism-related events using earables in a mimicked in-the-wild setting. Using inertial measurement unit for data collection, we utilise traditional machine learning for teeth grinding and clenching detection. We observe superior performance of models based on gyroscope data, achieving an 88% and 66% accuracy on grinding and clenching activities, respectively, in a controlled environment, and 76% and 73% on grinding and clenching, respectively, in an in-the-wild environment.

CCS CONCEPTS

• **Applied computing** → **Consumer health**; • **Computing methodologies** → *Classification and regression trees.*

KEYWORDS

earables; teeth grinding; bruxism; machine learning

ACM Reference Format:
Erika Bondareva, Elín Rós Hauksdóttir, and Cecilia Mascolo. 2021. Earables for Detection of Bruxism: a Feasibility Study. In *EarComp '21: 2nd International Workshop on Earable Computing, Sep 25, 2021, online.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

In the recent years, wireless earphones with built-in sensors, a.k.a. earables, have been gaining popularity — earphones are a commodity item providing established functionality, with support for privacy-preserving interaction by allowing the users to access information hands-free in a socially acceptable way, and, most importantly, have unique placement, allowing for numerous applications beyond playing music [10]. Comparing to other common areas of a human body for wearables placement (e.g. wrist), the ear is significantly more stationary, meaning that the collected signal is less susceptible to the external noise and motion artifacts, while also allowing to capture head and jaw movements, in addition to the activity in the rest of the body [10]. Due to these unique advantages offered by earable platforms, they are widely explored for health applications.

Drawing on the ability of wearables to capture jaw movements, we were interested in exploring the feasibility of using earables for detection of a movement disorder, characterised by teeth grinding and clenching, called bruxism. Bruxism affects around 8–13% of the adult population [2], and can cause numerous problems, affecting patient's teeth' health, causing headaches and disorders of the temporomandibular joint (TMJ). However, people suffering from bruxism often are unaware of the disorder until it becomes so advanced that the dentist is able to infer the diagnosis from the patient's worn down teeth. The exact cause of bruxism is unknown, but it is believed there is a genetic component to it [2], and it is usually linked to the levels of stress and anxiety that the patient is experiencing. With the COVID-19 pandemic having lasted for over a year and having had a major impact on the lives of nearly every individual, dentists are warning of another, accompanying, bruxism epidemic [7], prompting the issue of tooth wear and other side effects of bruxism.

Existing methods for diagnosing bruxism tend to be unreliable or invasive [18]. Most cases that are detected in earlier stages are based on self-reporting: for example, when a sleep partner notices grinding sounds, or when the patient reports TMJ pain. However, previous research indicates that the validity of the self-reported assessment of bruxism is low to modest and therefore is usually not sufficient for diagnosis of bruxism [21]. The golden standard for definitive diagnosis of bruxism is electromyogram (EMG) of the masticatory muscles by polysomnography audio-visual (PSG-AV) recording [4, 16, 19]. This method is performed in a controlled environment, and due to its complexity and high cost PSG-AV is not used for the assessment of bruxism in daily clinical practice.

To address the necessity of detecting bruxism in a non-invasive and low-cost way, this research presents a methodology to detect teeth clenching and grinding through an earable device by using traditional machine learning approaches. We collected accelerometer and gyroscope data from 17 participants using eSense wireless earbuds with a built-in inertial measurement unit (IMU), using the data from 13 participants for machine learning. The data comprised of participants grinding and clenching their teeth in a controlled environment, as well as performing bruxism-mimicking actions while engaging in routine activities that would simulate in-the-wild deployment. Namely, these activities included head movement, listening to music, walking, talking, chewing, and drinking.

[*]Both authors contributed equally to this research.

EarComp '21, Sep 25, 2021, Online

ACM ISBN 978-1-4503-XXXX-X/21/09…$15.00
https://doi.org/10.1145/1122445.1122456

To increase the inference accuracy, we collected the signals from both ears of each participant, and we extracted both time and frequency domain features. After preprocessing the collected data, we used traditional machine learning methods to develop a bruxism detection algorithm, using random forest (RF) and support vector machine (SVM). The detection algorithms were evaluated using both clean data and the more realistic signals collected with participants engaging in routine behaviours that could potentially affect the classification of teeth grinding events.

We show that by utilising traditional machine learning methods we can detect teeth grinding in a controlled environment with accuracy up to 88%. Additionally, we can detect teeth grinding with up to 76% accuracy for datasets mimicking real-world scenarios. The performance on clenching detection task is poorer, although still shows promise, with us achieving up to 73% accuracy on in-the-wild setting.

The main contributions of this work are as follows:

- We present a novel dataset compiled as a part of this study, which contains teeth grinding and clenching data collected through earables from 13 participants, in both noise-free environments and while performing routine activities to mimic in-the-wild data.
- We evaluate traditional signal processing and machine learning techniques for development of teeth grinding and clenching detection algorithms.
- We show the potential of detecting bruxism through earables by achieving 76% accuracy on in-the-wild teeth grinding and 73% accuracy on in-the-wild teeth clenching, and provide ideas for future directions.

2 RELATED WORK

EMG is a technology commonly utilised for a non-trivial task of bruxism diagnosis, capable of detecting mastication muscle movement. However, according to [9] EMG may detect muscle movement that is not necessarily related to bruxism, limiting the accuracy of methods relying solely on EMG. In addition, accuracy of portable EMG recorders for bruxism detection was reported as being unsatisfactory. [5] EMG can also be seen as obtrusive, relying on electrodes placed on the face for data collection. Therefore, it is important to explore alternative sensing modalities for detection of bruxism-related events.

There have been numerous efforts in the earables field exploring the potential of in-ear wearables for detection of activities related to the mouth.

A number of studies looked at detecting jaw and mouth movements by using earables. CanalSense presented a jaw, face, and head movement recognition system based on detecting changes in the air pressure inside the ear canal, using barometers embedded in earphones [1]. Another system, EarSense, sensed teeth gestures by detecting vibrations in the jaw that propagate through the skull to the ear, creating oscillations in the earphone speaker's diaphragm [17]. Other works looked at developing a separate system, rather than utilising earbuds, for detection of jaw movements. One such example would be the Outer Ear Interface that measured the deformation in the ear canal using proximity sensors caused by the lower jaw movements [3].

There have also been successful attempts at forming a human-computer interaction system based on unvoiced jaw movement tracking. JawSense considered the neurological and anatomical structure of the human jaw and cheek upon system design, and achieved successful classification of nine phonemes based on the muscle deformation and vibrations caused by unvoiced speaking [11].

[18] looked at using gyroscope data from an in-ear wearable for jaw clenching, which is an important part of what we set out to achieve in this feasibility study. The reported results had an error rate of 1% when the participant was seated and 4% when the participant moved, but the work was based on a single participant, and did not explore the detection of grinding.

Multiple works explored detection of bruxism using wearable devices, however, most of them are not as inconspicuous as a pair of earbuds. [13] introduced comfortable around the ear sensors (cEEGrids) for detecting awake bruxism, analysing bruxism-related events in contrast to the other facial activity events, such as chewing and speaking. [12] developed a wearable mouthguard with a force sensor to analyse teeth clenching during exercise. [6] proposed to detect sleep bruxism by using electromyography (EMG) and electrocardiogram (ECG) signals in combination, which produced substantially better results than using only EMG. [8] developed a system consisting of an interrogator/reader and a passive sensor that could be used to record bruxism-related events by placing the system in a dental splint. Finally, [20] developed a bite guard designed to analyse bruxism, with the monitoring achieved through a novel pressure-sensitive polymer.

To the best of our knowledge, this is the first work that looks at using wireless earbuds with built-in IMUs for detection of grinding and clenching with the goal of diagnosing and tracking bruxism.

3 STUDY DESIGN

An in-ear multisensory stereo device eSense was used for data collection. Specifically, we collected three-axis accelerometer and three-axis gyroscope data from the built-in IMU.

To collect the aforementioned data a mobile application, called eSense Client, was used for connecting via Bluetooth to the eSense earbuds and collecting raw IMU data. Due to the COVID-19 pandemic, all of the experiments were carried out remotely. Therefore, for annotating the data collected from the earables, a timestamped video was recorded using Zoom [22]. The collected eSense data was labelled by matching it with the video recording's timestamp.

Worth noting, that the eSense earbuds [10, 15] contain IMU only in the left earbud. Therefore, we used two left earbuds from two pairs of eSense for data collection, to explore the variation in accuracy for two sides, as well as potentially use the data gathered from both right and left ears together. Indeed, we discovered that participants typically had a dominant chewing side, which resulted in mastication muscles on one side of the face being greater developed. This resulted in data from one of the ears being more valuable for correct classification of bruxism-related activities, but since the dominant chewing side varies for different people, we had to use data from both right and left ear together.

The study was approved by the ethics committee in the Department of Computer Science and Technology at the University of

Cambridge. Informed consent was collected from the study participants. We ensured that no identifiable information was collected, and deleted the videos after the IMU data was labelled. Upon consulting a dental health professional, we also excluded any participants who suffer from bruxism to avoid any further damage to their teeth, as well as compiled a short questionnaire aimed at identifying potential participants who might be unknowingly suffering from bruxism. The questionnaire was based on Shetty's et al. research [19], but amended in collaboration with a certificated dentist to suit the needs of our study. In addition to these precautions, we also included a compulsory set of simple jaw massage exercises typically used in TMJ physiotherapy, aimed at alleviating any tension in the TMJ that the experiment might have inadvertently caused.

For the data collection, in total 17 participants were recruited, 12 females and 5 males, with the youngest participant aged 23 and the oldest aged 61. After verifying the quality of the collected data, data from four participants were discarded due to being compromised during data collection. Therefore, data from 13 participants were used for this research.

The aim of this study was to assess the feasibility of using in-ear wearables for detection of bruxism-related events, specifically teeth grinding and clenching. To address this goal, participants were asked to perform the following seven human-centred sensing experiments (with experiments 1-6 conducted with the participant in a sitting position):

1. Control experiment:

(a) Grind teeth for 5 seconds, pause for 5 seconds, repeat 6 times.
(b) Clench teeth for 5 seconds, pause for 5 seconds, repeat 6 times.

2. Moving head side to side:

(a) Look right and left for 30 seconds.
(b) Look right and left for the duration of 5 seconds while grinding, pause the grinding for 5 seconds, and repeat 6 times.
(c) Look right and left for the duration of 5 seconds while clenching, pause the clenching for 5 seconds, and repeat 6 times.

3. Chewing:

(a) Eat half a slice of bread.
(b) Chew gum for 30 seconds.

4. Read the provided text out loud for 30 seconds.
5. Drink 250 ml of water.
6. Listening to music:

(a) Listen to music for 30 seconds.
(b) While listening to music, grind teeth for 5 seconds, then pause for 5 seconds, and repeat 6 times.
(c) While listening to music, clench teeth for 5 seconds, then pause for 5 seconds, and repeat 6 times.

7. Walk around in a quiet room:

(a) Walk around in a quiet room for 30 seconds.
(b) Walk around for the duration of 5 seconds while clenching, pause the clenching for 5 seconds, and repeat 6 times.

The experiment took around 35 minutes for each participant to complete from start to finish.

To the best of our knowledge, this is the most advanced and sophisticated dataset that exists for in-ear IMU data of teeth grinding and clenching.

The experiments were designed each with a specific purpose in mind. While the first experiment was intended as control in a quiet, albeit unrealistic environment, the rest of the experiments were designed to either recreate activities that are known to interfere with signals collected via earbuds, such as moving the head, walking, as well as listening to music – an especially important task, keeping in mind that the primary purpose of earbuds is playing music. Other activities, such as chewing, drinking, and reading, were meant to recreate the activities in which a typical user is likely to engage daily, which include a significant involvement of mastication muscles, which might result in a signal similar to either clenching or grinding.

4 METHODOLOGY

4.1 Data collection

IMU (accelerometer and gyroscope) data was collected from both ears of the participants with a sampling rate of 5 Hz. This yielded approximately 25 minutes of IMU data per participant, containing 12 columns: X, Y, and Z axes for the acceleration and three axes for the gyroscope data collected from one ear, and the same data collected from the second ear. In addition to the IMU data, we also recorded a video of the participant performing the tasks, to use as ground-truth for correlating the collected IMU data to grinding and clenching events. Specifically, the videos were used to note down the start and end times of each general activity (such as moving head, chewing, walking, etc.), and also to note down the start and end times of grinding and clenching events during these activities. The times with no grinding or clenching events were noted down as silent periods, regardless of the general activity performed.

4.2 Segmentation and Labelling

Dealing with data in time domain, segmentation into shorter windows was necessary. We used a sliding window with 1.6 s length and 50% (or 0.8 s) overlap, to minimise the risk of missing the transition from one event into another.

Creating a label for each window posed a challenge due to the fact that sometimes the window would contain a transition from one event to another, such as transitioning from silent to grinding, making the labelling non-trivial. We explored two labelling methodologies:

- labelling the window as the dominant event in that window: if the larger portion of the window contains the grinding data and smaller portion of silent data, the window would be labelled as grinding. If the split is equal, then the window is labelled as silent.
- only labelling the window as silent if all the samples within the window are silent, and no amount of grinding or clenching event present.

Based on the preliminary comparison of the methodologies, it appeared that labelling the window according to the dominant event yielded superior performance, due to which this was the method that we chose for further analysis.

4.3 Feature Extraction

As input for the machine learning algorithms we used a number of various features. Raw signal from both ears was used, resulting in 8 datapoints for each axis for each ear, yielding 48 features. Then, we used a sum of vector magnitudes (SOVM), which was calculated using the following equation:

$$\mathrm{SOVM} = \sqrt{x_R^2 + y_R^2 + z_R^2} + \sqrt{x_L^2 + y_L^2 + z_L^2},$$

where x, y, and z represent a single value collected from the sensor, corresponding to one of the axes, and L and R representing the signals collected from left and right ear, respectively.

The raw data collected from the three axes and the calculated SOVM can be seen in Figure 1.

We utilised commonly-used Python libraries Librosa and Scipy to extract a few additional features: Mel Frequency Cepstral Coefficients (MFCCs), spectral flatness, spectral centroid, and poly features calculated for the sum of vector magnitudes, and a mean was calculated for each of the additional feature vectors. Then, maximum, minimum, mean, standard deviation, and absolute deviation of the signal amplitude was extracted from the sum of vector magnitudes and concatenated with the rest of the features. Finally, zero-crossing rate for each of the axes was averaged and concatenated. This yielded a total of 71 features.

4.4 Classification

Five traditional machine learning classifiers were explored: decision tree (DT), k-nearest neighbours (k-NN), logistic regression (LR), random forest (RF), and support vector machine (SVM). In this work, the results for RF and SVM are reported, since they demonstrated the best performance during preliminary evaluation.

For evaluating the detection algorithm, we used a leave-one-out approach, commonly utilised in machine learning problems with limited number of participants. Then, mean and standard deviation were calculated for each of the performance metrics across the resulting values.

To evaluate the algorithm, we compared accuracy, precision, recall, and f1-score. While low number of false positives (FP) and false negatives (FN) is desirable, it is important to keep in mind that in this scenario occasional FP or FN might be acceptable due to the ability to infer the overall diagnosis only if multiple bruxism-related events are detected.

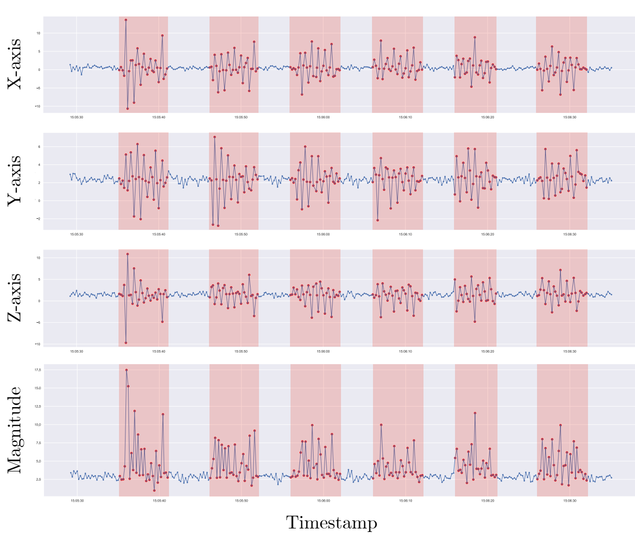


Figure 1: Raw gyroscope data collected with the participant grinding teeth, with the grinding events coloured in red, and periods with no grinding coloured in grey.

4.5 Experiment design

The goal of this study was to assess the feasibility of detecting bruxism-related events in a controlled and a mimicked in-the-wild environment through in-ear wearables using IMU signals. This goal informed our experiment design. In this paper we present the performance of two traditional machine learning algorithms, RF and SVM, on data collected from two different sensors, accelerometer and gyroscope, for the following tasks:

Task 1: detection of bruxism-related events in a controlled environment, with minimal external noise and no other actions performed, participant being still:

(a) for teeth grinding (580:388 windows of grinding:silent data);
(b) for teeth clenching (589:438 windows of clenching:silent data).

Task 2: detection of bruxism-related events in a mimicked in-the-wild environment (in order to evaluate algorithm performance in a more realistic setting), with participants performing a range of routine activities:

(a) teeth grinding and clenching during no general activity, while moving head, and while listening to music (1658:2334 grinding:silent and 1695:2334 clenching:silent windows);
(b) activities from task 2b with the addition of other routine activities, such as chewing bread, chewing gum, reading out loud (to imitate speaking), drinking water, and walking. Worth noting, that the additional activities do not include grinding or clenching events, and are intended for testing the detection algorithm performance on actions that are known to either be relatively noisy or involve significant jaw activity that may be misclassified as teeth grinding or clenching (1658:4825 grinding:silent and 1695:4815 clenching:silent windows).

5 RESULTS AND DISCUSSION

5.1 Bruxism Detection in Controlled Environment

Task 1 was designed with the purpose of assessing the feasibility of detecting teeth grinding and clenching using earbuds with a built-in IMU in a controlled environment, with the participant completely still. Based on leave-one-out validation, comparing SVM to RF demonstrates that SVM performs better on both grinding and clenching detection. It is obvious that using gyroscope data yields a significantly higher performance, achieving 89% and 60% on grinding and clenching events detection, respectively. For the controlled experiment, RF achieves superior performance only on detection of grinding events using accelerometer data (70% accuracy), in comparison to detection of grinding events using accelerometer and SVM (66% accuracy). However, both results are significantly worse than the detection accuracy achieved using gyroscope data. We also calculated precision, recall, and f-1 score, and these metrics as well as the standard deviation for the leave-one-out cross-validation are reported in Table 1.

5.2 Bruxism Detection In-The-Wild

We demonstrated that in-ear wearables show promise for detection of bruxism-related events in a controlled environment. But, naturally, it is important to also analyse whether earables could offer a viable solution for unobtrusive bruxism detection in-the-wild. For this purpose, Task2(a) and Task2(b) were designed to test the bruxism-related activity detection approach while the study participants were performing other actions. Specifically, Task2(a) focused on teeth grinding and clenching in silence, as well as these events while listening to music or moving the head. Task2(b) presented an even more complex problem, adding other activities that involve substantial jaw movement.

For Task2(a), SVM model on gyroscope data showed the best performance, for teeth grinding detection yielding 73% accuracy, and reaching 61% accuracy for teeth clenching detection. Task2(b) proved to be a more challenging experiment, which can be observed from reduced performance on precision, recall, and f-1 score. However, RF still performs sufficiently well on gyroscope data, yielding a 68% and 61% precision on grinding and clenching, respectively. SVM was incapable of detecting clenching using accelerometer data, predicting that all the testing data do not contain clenching instances. Therefore, no metrics are reported for this algorithm. In general, acceleration-based performance was insufficient, proving that gyroscope provides more valuable data for bruxism detection. Detailed results for these experiments can be seen in Tables 2 and 3.

We also evaluated which tasks have the most impact on detection of bruxism, concluding that head movements and reading out loud have the most impact on correct classification of grinding events, and head movements, walking, and drinking have the most impact on clenching detection.

Table 1: Detection of grinding (denoted as Gr.) and clenching (denoted as Cl.) by SVM and RF algorithms on Task 1. The values reported are mean±stdev.

		Gyroscope		Accelerometer	
		SVM	RF	SVM	RF
Gr.	Accuracy	0.88±0.05	0.85±0.08	0.66±0.12	0.70±0.11
	Precision	0.89±0.04	0.87±0.07	0.69±0.19	0.70±0.16
	Recall	0.88±0.05	0.84±0.09	0.63±0.12	0.67±0.12
	f1-score	0.88±0.05	0.83±0.10	0.57±0.17	0.63±0.15
Cl.	Accuracy	0.60±0.06	0.57±0.06	0.58±0.05	0.55±0.06
	Precision	0.56±0.13	0.54±0.10	0.40±0.15	0.53±0.12
	Recall	0.57±0.06	0.54±0.07	0.51±0.02	0.52±0.08
	f1-score	0.52±0.11	0.52±0.11	0.40±0.06	0.49±0.08

Table 2: Performance on detection of teeth grinding and clenching on Task2(a).

		Gyroscope		Accelerometer	
		SVM	RF	SVM	RF
Gr.	Accuracy	0.73±0.05	0.73±0.08	0.55±0.04	0.56±0.07
	Precision	0.74±0.06	0.73±0.08	0.43±0.10	0.51±0.12
	Recall	0.70±0.05	0.71±0.08	0.49±0.03	0.52±0.07
	f1-score	0.70±0.06	0.71±0.09	0.42±0.06	0.48±0.09
Cl.	Accuracy	0.61±0.05	0.60±0.04	0.57±0.03	0.52±0.06
	Precision	0.60±0.08	0.57±0.07	0.36±0.15	0.45±0.08
	Recall	0.56±0.06	0.57±0.05	0.50±0.01	0.49±0.02
	f1-score	0.52±0.08	0.55±0.07	0.38±0.04	0.42±0.04

Table 3: Performance on detection of teeth grinding and clenching on Task2(b).

		Gyroscope		Accelerometer	
		SVM	RF	SVM	RF
Gr.	Accuracy	0.74±0.03	0.76±0.05	0.73±0.03	0.67±0.11
	Precision	0.53±0.18	0.68±0.07	0.46±0.14	0.48±0.10
	Recall	0.51±0.02	0.61±0.06	0.50±0.02	0.49±0.05
	f1-score	0.45±0.04	0.61±0.07	0.44±0.04	0.45±0.07
Cl.	Accuracy	0.74±0.02	0.73±0.02	N/A	0.70±0.04
	Precision	0.39±0.05	0.61±0.05		0.41±0.05
	Recall	0.50±0.00	0.54±0.02		0.49±0.02
	f1-score	0.43±0.01	0.52±0.03		0.44±0.03

5.3 Limitations and Future Work

Although this work is intended as a feasibility study to assess the capability of in-ear wearables to detect bruxism-related events, it is nevertheless important to highlight its limitations and discuss some potential avenues worth exploring in the future.

Teeth grinding, which is the typical symptom of bruxism, usually happens when the person is not fully conscious, most often exhibited in patients' sleep. Our dataset consists of people who are unlikely to be actually suffering from bruxism, and the data is collected with the participants being awake and fully conscious, which means that the data which would be collected from real bruxism sufferers might be slightly different. However, earbuds for sleep have started to appear on the market, the sole purpose of which is to provide comfortable noise cancellation, which is very promising for unobtrusive detection of bruxism during sleep.

Our results show that the accuracy of clenching detection is lower than that of grinding. However, it is important to note that grinding is an action that involves continuous movement in the jaw joint, resulting in continuously changing IMU data. Clenching, on the other hand, only involves changes in acceleration and angular velocity when the action is initiated, i.e. the person clasps their teeth together, and not when the action is in progress, i.e. the teeth are clasped and the jaw is not moving. This may explain the poorer performance on detection of clenching events. Going forward, exploring alternative ways of data segmentation and labelling might be useful, as well as more advanced machine learning approaches that are capable of dealing with time-series data, such as recurrent neural networks (RNNs).

From the classification perspective, given recent advances in deep learning, it would be imperative to explore the potential of deep learning on raw IMU data from earables for detection of bruxism-related events.

Other potential areas of research would include exploring sensor fusion and considering accelerometer and gyroscope data in combination, as well as potentially investigating the feasibility of using audio collected with a microphone pointing inside the ear [14].

Finally, given the complexity of collecting and labelling bruxism data, it would be interesting to explore IMU data augmentation, as well as potentially generating synthetic data.

6 CONCLUSIONS

This work presents a feasibility study on detection of bruxism-related events using earables. During a bruxism epidemic, which was potentially exacerbated by the COVID-19 pandemic, it is essential to devise a low-cost, unobtrusive, and socially acceptable method for bruxism detection, which would allow to diagnose the patients in early stages of the disorder, before irreversible tooth wear occurs. We compiled a first extensive dataset of teeth grinding and clenching data collected via earbuds with a built-in IMU. In addition to collecting these data in a controlled environment, we also collected data mimicking in-the-wild signal by asking the users to simulate teeth grinding and clenching while performing other activities, or to engage in routine activities that require substantial jaw involvement.

By using traditional machine learning methods, we concluded that SVM and RF yield the best performance. Gyroscope data appears to be much more valuable than acceleration data for identification of bruxism-related events. We achieved 88% and 66% accuracy on teeth grinding and clenching, respectively, in a controlled environment. We also demonstrate the potential of this technology in a mimicked in-the-wild environment, achieving 76% and 73% accuracy on teeth grinding and clenching, respectively.

ACKNOWLEDGMENTS

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/L015889/1 for the Centre for Doctoral Training in Sensor Technologies and Applications, ERC Project 833296 (EAR), and Nokia Bell Labs through their donation for the Centre of Mobile, Wearable Systems and Augmented Intelligence.

REFERENCES

[1] Toshiyuki Ando, Yuki Kubo, Buntarou Shizuki, and Shin Takahashi. 2017. CanalSense: Face-Related Movement Recognition System Based on Sensing Air Pressure in Ear Canals. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, 679–689. https://doi.org/10.1145/3126594.3126649

[2] H. Beddis, M. Pemberton, and Stephen Davies. 2018. Sleep Bruxism: an Overview for Clinicians. *British Dental Journal* 225, 6 (2018), 497–501. https://doi.org/10.1038/sj.bdj.2018.757

[3] Abdelkareem Bedri, David Byrd, Peter Presti, Himanshu Sahni, Zehua Gue, and Thad Starner. 2015. Stick It in Your Ear: Building an in-Ear Jaw Movement Sensor *(UbiComp/ISWC'15 Adjunct)*. Association for Computing Machinery, 1333–1338. https://doi.org/10.1145/2800835.2807933

[4] E. Casett, J. C. Réus, J. Stuginski-Barbosa, A. Porporatti, M. Carra, M. Peres, G. De Luca Canto, and D. Manfredini. 2017. Validity of different tools to assess sleep bruxism: a meta-analysis. *Journal of Oral Rehabilitation* 44 (2017), 722–734.

[5] T. Castroflorio, A. Bargellini, G. Rossini, G. Cugliari, A. Deregibus, and D. Manfredini. 2015. Agreement between clinical and portable EMG/ECG diagnosis of sleep bruxism. *Journal of Oral Rehabilitation* 42, 10 (2015), 759–764.

[6] Tommaso Castroflorio, Luca Mesin, Gianluca Martino Tartaglia, Chiarella Sforza, and Dario Farina. 2013. Use of Electromyographic and Electrocardiographic Signals to Detect Sleep Bruxism Episodes in a Natural Environment. *IEEE Journal of Biomedical and Health Informatics* 17, 6 (2013), 994–1001. https://doi.org/10.1109/JBHI.2013.2274532

[7] Dadnam D., Dadnam C. Al-Saffar H. 2021. Pandemic Bruxism | British Dental Journal 230, 271. https://www.nature.com/articles/s41415-021-2788-3#citeas.

[8] C. González and A Díaz Lantada. 2009. A wearable passive force sensor powered by an active interrogator intended for intra-splint use for the detection and recording of bruxism. In *2009 3rd International Conference on Pervasive Computing Technologies for Healthcare*. 1–4. https://doi.org/10.4108/ICST.PERVASIVEHEALTH2009.5884

[9] Nantawachara Jirakittayakorn and Yodchanan Wongsawat. 2014. An EMG instrument designed for bruxism detection on masseter muscle. In *The 7th 2014 Biomedical Engineering International Conference*. 1–5.

[10] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Alessandro Montanari. 2018. Earables for Personal-Scale Behavior Analytics. *IEEE Pervasive Computing* 17, 3 (2018), 83–89. https://doi.org/10.1109/MPRV.2018.03367740

[11] Prerna Khanna, Tanmay Srivastava, Shijia Pan, Shubham Jain, and Phuc Nguyen. 2021. JawSense: Recognizing Unvoiced Sound Using a Low-Cost Ear-Worn System. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications* (Virtual, United Kingdom) *(HotMobile '21)*. Association for Computing Machinery, 44–49. https://doi.org/10.1145/3446382.3448363

[12] Rio Kinjo, Takahiro Wada, Hiroshi Churei, Takehiro Ohmi, Kairi Hayashi, Kazuyoshi Yagishita, Motohiro Uo, and Toshiaki Ueno. 2021. Development of a Wearable Mouth Guard Device for Monitoring Teeth Clenching during Exercise. *Sensors* 21, 4 (2021). https://doi.org/10.3390/s21041503

[13] Michael Knierim, Max Schemmer, and Dominik Woehler. 2021. Detecting Daytime Bruxism Through Convenient and Wearable Around-the-Ear Electrodes. In *12th International Conference on Applied Human Factors and Ergonomics (AHFE 2021)*.

[14] 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* (Virtual Event, Wisconsin) *(MobiSys '21)*. Association for Computing Machinery, 175–187. https://doi.org/10.1145/3458864.3467680

[15] Chulhong Min, Akhil Mathur, and Fahim Kawsar. 2018. Exploring Audio and Kinetic Sensing on Earable Devices. In *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications* (Munich, Germany) *(WearSys '18)*. Association for Computing Machinery, 5–10. https://doi.org/10.1145/3211960.3211970

[16] Lucas B Pigozzi, Daniela D S Rehm, Simone C Fagondes, Eduardo P Pellizzer, and Márcio L Grossi. 2019. Current Methods of Bruxism Diagnosis: A Short Communication. *The International journal of prosthodontics* 32, 3 (2019), 263—264. https://doi.org/10.11607/ijp.6196

[17] Jay Prakash, Zhijian Yang, Yu-Lin Wei, Haitham Hassanieh, and Romit Roy Choudhury. 2020. EarSense: Earphones as a Teeth Activity Sensor. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking* (London, United Kingdom) *(MobiCom '20)*. Association for Computing Machinery, Article 40. https://doi.org/10.1145/3372224.3419197

[18] Siddharth Rupavatharam and Marco Gruteser. 2019. Towards In-Ear Inertial Jaw Clenching Detection. In *Proceedings of the 1st International Workshop on Earable Computing* (London, United Kingdom) *(EarComp'19)*. Association for Computing Machinery, 54–55. https://doi.org/10.1145/3345615.3361134

[19] Shetty, S., Pitti, V., C.L., Babu, S., G.P., Kumar, S. and Deepthi, B. C. . 2010. Bruxism: A Literature Review. J Indian Prosthodont Soc 10, 141–148 (2010). https://link.springer.com/article/10.1007/s13191-011-0041-5.

[20] M. van Dijk. 1995. The binary symmetric broadcast channel with confidential messages, with tampering. In *Proceedings of 1995 IEEE International Symposium on Information Theory*. 487–. https://doi.org/10.1109/ISIT.1995.550474

[21] Wataru Yachida, Taro Arima, Eduardo E. Castrillon, Lene Baad-Hansen, Noboru Ohata, and Peter Svensson. 2016. Diagnostic validity of self-reported measures of sleep bruxism using an ambulatory single-channel EMG device. *Journal of Prosthodontic Research* 60, 4 (2016), 250–257. https://doi.org/10.1016/j.jpor.2016.01.001

[22] Zoom Video Communications Inc. 2016. Zoom Video Communications Inc . (2016). Security guide. Zoom Video Communications Inc. https://d24cgw3uvb9a9h.cloudfront.net/static/81625/doc/Zoom-Security-White-Paper.pdf.

Detecting forward leaning posture using eSense and developing a posture improvement promoting system

Yushi Takayama
Aoyama Gakuin University
Tokyo, Japan
ytakayama@wil-aoyama.jp

Shun Ishii
Aoyama Gakuin University
Tokyo, Japan
sishii@wil-aoyama.jp

Anna Yokokubo
Aoyama Gakuin University
Tokyo, Japan
yokokubo@it.aoyama.ac.jp

Guillaume Lopez
Aoyama Gakuin University
Tokyo, Japan
guillaume@it.aoyama.ac.jp

ABSTRACT

In this study, we propose a method using eSense for detecting forward leaning posture during computer work, and verify its effectiveness. We will also develop a system that prompts the user to improve posture when a forward leaning posture is detected, and verify its effectiveness. In the future, we aim to develop a system that targets not only forward leaning posture but also posture improvement in other parts of the body by utilizing the contents verified in this research.

CCS CONCEPTS

• **Computer systems organization → Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

earable device, wearable device, posture improvement, office work

ACM Reference Format:
Yushi Takayama, Shun Ishii, Anna Yokokubo, and Guillaume Lopez. 2021. Detecting forward leaning posture using eSense and developing a posture improvement promoting system. In *EarComp 2021: 2nd International Workshop on Earable Computing in Conjunction with Ubicomp 2021, September 25–26, 2021.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Office workers, who work with a computer, often suffer from unnatural posture while working. Working for a long time with unnatural posture can cause not only physical fatigue, but also diseases such as tendinitis and hunchback. In order to prevent the onset of these diseases, it is necessary to work with correct posture continuously. However, it seems difficult to work while always being aware of maintaining correct posture. We think it is necessary that a system that detects unnatural posture and notify users to encourage them to maintain correct posture. In this study, we focus on detecting forward leaning posture, and aim to verify the method of it using eSense, which is an earable device equipped with a 6-axis acceleration sensor. At the same time, we aim to develop a system that encourages posture improvement when a forward leaning posture is detected, and to verify the effectiveness of the system.

2 RELATED WORK

Several posture estimation methods using wearable devices have been proposed in previous studies. For example, this study[2] verified whether posture estimation is possible using only the 3-axis accelerometer of a cell phone, and found that the basic motions of sitting, standing, walking, and running could be estimated. However, in order to detect forward leaning posture using a cell phone, it needs to be worn on the upper body where body movements are large. Since cell phones have been getting larger and larger on recent years, wearing them on the upper body all the time can be a hindrance to your work. Also, not all people wear a cell phone while working with a computer (for example, some people put it on their desk while working).

On the other hand, there is also previous study about a system that promote improved posture during desk work. For example, this study[3] investigated the use of a very slow moving monitor for unobtrusive posture correction, and revealed its effectiveness. We think this system very good in that it can correct posture imperceptibly without disturbing the user's work with sound or vibration, but the drawback is that it requires a motorized monitor which is hardly widespread yet.

The device used for posture estimation while working on a computer should not interfere with the work and should not cause discomfort even if it is worn for a long time. Therefore, we think earphones are appropriate because it is small and familiar to many people. Although sensor-equipped earphones are not yet widely available, given the popularity of TWS over the past few years, we think it would be useful to conduct this study using earables.

3 METHODS

3.1 Detecting forward leaning posture

In this study, we use eSense, which is an earable device developed by Nokia Bell Labs[1]. eSense is equipped with a 6-axis accelerometer

EarComp 2021, September 25–26, 2021,

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

in the left earbud, and we propose a method to detect forward-leaning posture from these sensor data (see Fig. 1). To determine whether the posture is forward-leaning or not, we can use the rotation angle of the left earbud. As shown in Figure 2, when users are in forward-leaning posture, there should be a change in the rotation angle on the Z-axis. Therefore, we detect forward-leaning posture when the a rotation angle on the Z-axis of the left earbud exceeds a threshold value for a certain amount of time.

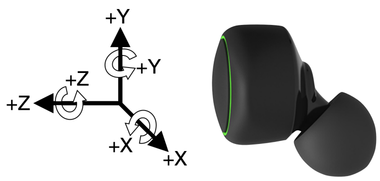


Figure 1: The orientation of the IMU axes and the polarity of rotation in relation to the earbud (quoted from [1]).

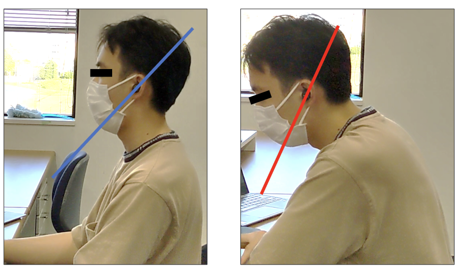


Figure 2: There should be a change in the rotation angle on the Z axis when users are in forward leaning posture.

3.2 Implementing a posture improvement promoting system

Since we can access the 6-axis accelerometer data through the BLE interface, we created an application to acquire the sensor values in real-time. To improve focus on the work or studies, it is recommended to keep smartphones out of sight. However, it is not the case for smartwatches, and smartwatches can also be used to collect further vital data that reflects user's state, such as hand motion and heart rate. Hence, we propose to collect eSense data from a BLE enable Android OS based smartwatch. Android API used to provide two ways for measuring the rotation angle of android devices, but $Sensor.TYPE\_ORIENTATION$ has already been deprecated, and $SensorManager.getOrientation()$ cannot be used with eSense because it requires the value of a magnetometer, which eSense is not equipped. Therefore, we need to implement the algorithm to measure the rotation angle of the device by ourselves. We consider using the following two methods to measure the rotation angle and will verify their accuracy.

(a) Calculate the rotation angle by time integration of angular velocity, and correct it using the acceleration data.
(b) Calculate the rotation angle by applying a Kalman filter to acceleration data and angular velocity data.

Then, we propose three kinds of feedback methods to encourage users to improve their posture.

- The first method is to play an audio message from the eSense. It is inspired by the fact that people will wear earbuds while using this system, and it is expected to validate the effectiveness of auditory feedback while working.
- The second method is to show notifications on the computer screen. By using the OS notifications that users often see while working, it is expected to provide feedback without feeling users disturbed their work.
- The third method is to activate vibration and show notification on the smartwatch.

Figure 3 show a schematic of proposed system. We plan to evaluate the effectiveness of each method by asking the subjects to fill out a questionnaire and assess whether it helped improve their posture and whether it did not interfere with their work.

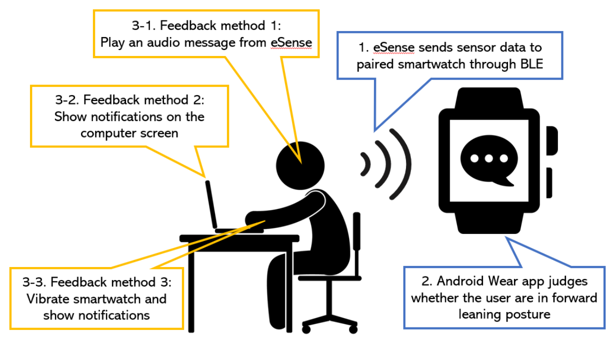


Figure 3: Schematic overview of proposed posture improvement system

4 FUTURE WORK

Both eSense and smartwatches do not interfere with work and do not cause discomfort even when worn for a long time. Besides measuring the forward-leaning posture from eSense, measuring the wrist tilt angle using the smartwatch and providing feedback to users may help prevent tendinitis caused by prolonged typing. Many smartwatches do not have a geomagnetic sensor but have an accelerometer and gyroscope. Hence, the method of measuring the rotation angle proposed in this study may be applicable.

In this way, we would like to develop a posture improvement promoting system that combines multiple daily usage devices and lead to the verification of its effectiveness.

REFERENCES

[1] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Alessandro Montanari. 2018. Earables for Personal-Scale Behavior Analytics. *IEEE Pervasive Computing* 17, 3 (2018), 83–89. https://doi.org/10.1109/MPRV.2018.03367740

[2] Takashi Morito, Toshiki Hayashi, and Hiroyuki Morikawa. 2006. User Posture and Movement Estimation Based on 3-Axis Acceleration with a Sensor Network. In *The Special Interest Group Technical Reports of IPSJ.*

[3] Joon-Gi Shin, Eiji Onchi, Maria Jose Reyes, Junbong Song, Uichin Lee, Seung-Hee Lee, and Daniel Saakes. 2019. *Slow Robots for Unobtrusive Posture Correction.* Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3290605.3300843

Earable Design Analysis for Sleep EEG Measurements

Swati Mandekar
IfB
FH Aachen
Aachen, Germany
swati.mandekar@alumni.fh-aachen.de

Lina Jentsch
D-HEST
ETH Zurich
Zurich, Switzerland
jentschl@ethz.ch

Dr. Kai Lutz
IDUN Technologies AG
Zurich, Switzerland
kai@iduntechnologies.ch

Dr. Mehdi Behbahani
IfB
FH Aachen
Aachen, Germany
behbahani@fh-aachen.de

Dr. Mark Melnykowycz
IDUN Technologies AG
Zurich, Switzerland
mark@iduntechnolgies.ch

ABSTRACT

Conventional EEG devices cannot be used in everyday life and hence, past decade research has been focused on Ear-EEG for mobile, at-home monitoring for various applications ranging from emotion detection to sleep monitoring. As the area available for electrode contact in the ear is limited, the electrode size and location play a vital role for an Ear-EEG system. In this investigation, we present a quantitative study of ear-electrodes with two electrode sizes at different locations in a wet and dry configuration. Electrode impedance scales inversely with size and ranges from 450 kΩ to 1.29 MΩ for dry and from 22 kΩ to 42 kΩ for wet contact at 10 Hz. For any size, the location in the ear canal with the lowest impedance is ELE (Left Ear Superior), presumably due to increased contact pressure caused by the outer-ear anatomy. The results can be used to optimize signal pickup and SNR for specific applications. We demonstrate this by recording sleep spindles during sleep onset with high quality (5.27 μVrms).

CCS CONCEPTS

• Human-centered computing • Ubiquitous and mobile computing • Ubiquitous and mobile computing systems and tools

KEYWORDS

EEG, sensors, BCI, Ear-EEG, biopotential electrodes, Sleep EEG, Impedance Spectroscopy

ACM Reference format:

Swati Mandekar, Lina Jentsch, Dr. Kai Lutz, Dr. Mehdi Behbahani and Dr. Mark Melnykowycz. 2021. I Earable Design Analysis for Sleep EEG Measurements. In *EarComp'21: Proceedings of the 2nd International Workshop on Earable Computing.*

1 Introduction

Electroencephalography (EEG) is a technique to measure electric potential differences produced by a group of neurons measured typically at the surface of the scalp. The EEG devices used in clinical settings are bulky, require trained personnel, and are time-consuming to setup and conduct measurements with. Hence, over the past decade, many research groups have investigated recording EEG from the outer ear canal (Ear-EEG) [1, 3, 4, 8, 14]. Ear-EEG has been shown to have similar features and quality to traditional scalp EEG [10]. The combination with a miniaturized and energy efficient biopotential front-end widened the scope to prospective uses outside of clinical and research areas. As a result, Ear-EEG appears to be the only viable choice for widespread consumer adoption of EEG technology. Applications for Ear-EEG range from brain-computer interfaces [11], work-related stress management solutions [7], and sleep monitoring [13] to cognitive-controlled hearing aids [2], epilepsy seizure warning devices [21], emotion monitoring [1] and neurofeedback home training systems [5].

As EEG records synchronized neuronal electrical activity, the signal needs to travel from the point of its origin i. e., clusters of neurons to the surface of the head through different tissue layers. For good signal acquisition, apart from the measurement device requirements to have a high input impedance and common mode rejection ratio (CMRR), the skin-electrode interface plays a crucial role. The skin contact impedance depends on the electrode's properties such as the electrode's material, size and skin conformity. Particularly for mobile applications, it is important to achieve a low-impedance contact to obtain a sufficiently high signal-to-noise ratio (SNR).

Kappel et Al. [9] investigated impedance inside the ear canal to compare wet and dry ear-EEG electrodes. They observed a mean (standard deviation) low frequency impedance (<100 Hz) of the canal electrodes of 1.2 MΩ (SD =1.4 MΩ) and a high frequency impedance (>100 Hz) of 230 kΩ (SD = 220 kΩ) for rectangular silver electrodes of area 14 mm². The impedance measurements accounting for cerumen showed an 86% decrease in impedance after removing cerumen [18]. The electrode stabilization time in the ear before a measurement is also a major determinant of stability and low skin contact impedance, lowering impedance by up to 50% [15]. Hence, in this research, all measurements were conducted after removing the cerumen from the ear canal and after specific stabilization times.

In this study, we extend the discussion by investigating the influence of electrode size and location on the skin contact impedance. We compare the impedance for two electrode sizes and three locations inside the ear canal for dry and wet contact. Based on the findings, the optimal size and position are used to measure and characterize Ear-EEG during sleep onset. Using the proposed configuration, we were able to successfully measure a high-quality signal and extract characteristic sleep features from the measured data.

The experiments and data produced were part of the research for the Ear-EEG acquisition system for IDUN Technologies AG.

2 Materials and Methods

2.1 Earpiece Production

A memory foam earplug is used as a base to provide more flexibility and comfort for long-duration monitoring and to adapt to the different ear canal shapes of the test subjects. The electrodes were fabricated by adding cutout silver-plated knitted conductive fabric (surface resistivity of below 5 Ω/cm^2) to the surface of the foam earplugs in order to match the mechanical deformation of the memory foam. Each earpiece comprises three electrodes oriented at an angle of 120° relative to the longitudinal axis as shown in figure 1. Two different sizes were laser cut and bonded to the earplugs using biocompatible adhesive at 120° apart as shown in Figure 1.

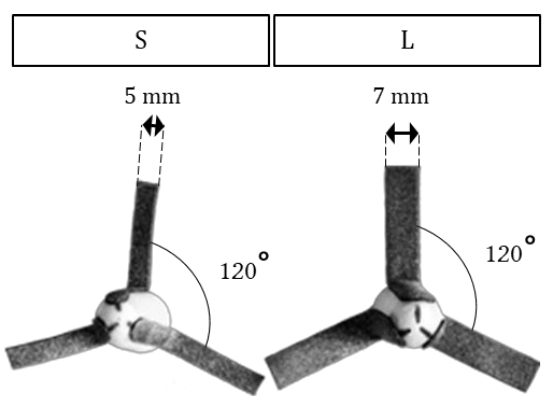


Figure 1. Foam earplug earpieces with three electrodes located at 120° apart of sizes 'Small' 8x5 mm (40 mm^2) and 'Large' 8x7 mm (56 mm^2)

Each electrode was manufactured with a curvature at the tip in order to avoid skin irritation while inserting in the ear canal.

2.2 Electrode Configuration

The labeling scheme for in-ear electrode locations was adapted from P. Kidmose et Al. [10]. The electrodes are indicated by Exy, where x ∈ {L, R} refers to electrodes in the left or right ear, and y refers to the electrode position in the ear canal respectively. The electrode locations investigated in this study were ELE, ELJ, and ELH as shown in Figure 2.

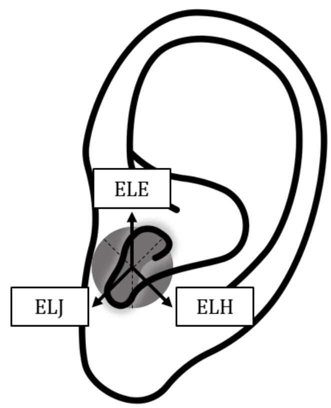


Figure 2. Labeling scheme for the in-ear electrodes, Exy where x ∈ {L, R} refers to the left or right ear and y is the position of the electrodes. The three positions used in this study are ELE, ELH and ELJ

2.3 Test Subjects

The measurements were performed in the left ear of 10 healthy participants after a stabilization time of 5-7 minutes. Before measurements, the participants were requested to clean their ears with a cotton swab and ethyl alcohol wipe before inserting earpieces in the ear. For dry measurements, electrodes were inserted into the cleaned but dry ear canal. For wet measurements, a drop of saline solution was applied on each electrode before the experiment. Before putting gel reference and counter electrodes on the skin, the skin was prepared by cleaning with ethyl alcohol. For hygienic concerns, each participant received their personal earpieces.

2.4 Impedance Spectroscopy

A frequency response analyzer (PalmSens4 [16]) was used to evaluate the frequency response in a three-wire impedance configuration which was connected to a Laptop (Windows Version 10) running PSTrace software [17].

The impedance measurements were carried out at the same time on all three electrodes simultaneously, whereas ELJ, ELH, and ELE served as working electrodes (WE) and standard wet gel electrodes placed at the inion as a counter (CE) and between inion and ear as reference (RE) electrodes as shown in Figure 3. Wires were crimped to the electrodes using connector pins. The medical tape was used to ensure that the electrodes outside of the ear canal did not contact each other. After the stabilization time, four frequencies were swept from 1 Hz to 100 Hz. The data is analyzed using MATLAB with mean and standard deviation.

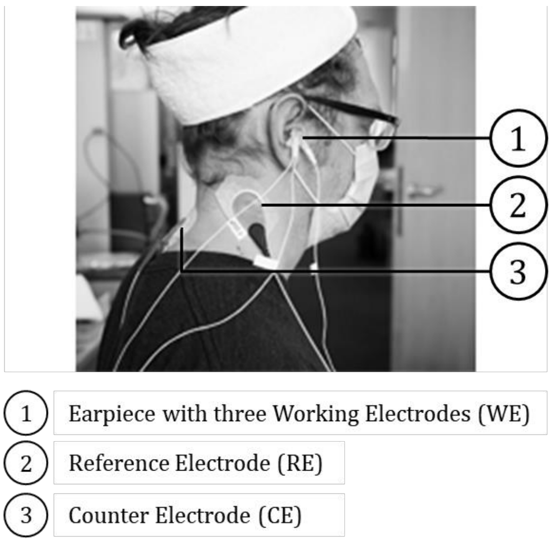


Figure 3. Electrode configuration of the ear impedance spectroscopy on the test subject for the three-electrode setup, standard wet gel electrodes are at RE and CE, WE are ear electrodes

2.5 Sleep EEG

As an initial investigation into the ability of the three electrode earpieces to detect EEG features, a 30-minute sleep session was organized. An Open BCI Cyton device was used to amplify biopotential signals from the left ear of the subject and were recorded using the OpenBCI Graphical User Interface (GUI) software package. The test subject reclined on a comfortable sofa and fell asleep while being monitored with a video camera to identify large body movements and aid in the results interpretation.

A stabilization time of ~10 minutes was used to allow the memory foam to stabilize in the ear. The only preparation done was cleaning ears with cotton swabs. The contra-lateral mastoid position was used for the counter and the neck was used for the reference position of the electrode setup. Standard wet gel electrodes were used as counter and reference electrodes.

3 Results and Discussion

3.1 Electrode Size

The mean impedance after 5 and 7 minutes of stabilization time in the dry ear for S and L electrode sizes is shown in Figure 4 (n=3). The impedance was inversely proportional to the size of the electrodes and ranged from 1.25 MΩ for small to 450 kΩ for large electrodes after 5 min of stabilization time and from 1.29 MΩ (S) to 511 kΩ (L) after 7 mins of stabilization time at 10 Hz as shown in Figure 4. As reported previously in the literature, intra- and inter-subject variability can be significant for dry contact electrodes. For visual clarity, the standard deviation was excluded from the plot but can be seen in Table 1 in the supporting material.

Depending on the used amplifier, impedances should typically be well below 1 MΩ to enable high-quality signal recordings and artifact rejection [6]. At the same time, bigger electrodes mean loss of spatial resolution and number of channels. To balance this tradeoff, we decided to use the smaller size for the further experiments but reduce the impedance by adding a drop of saline solution to the electrode surface prior to measurement.

In contrast to the electrode size, the two minutes difference in stabilization time did not significantly change the impedance; hence a stabilization duration of 5 minutes was chosen for the following experiments.

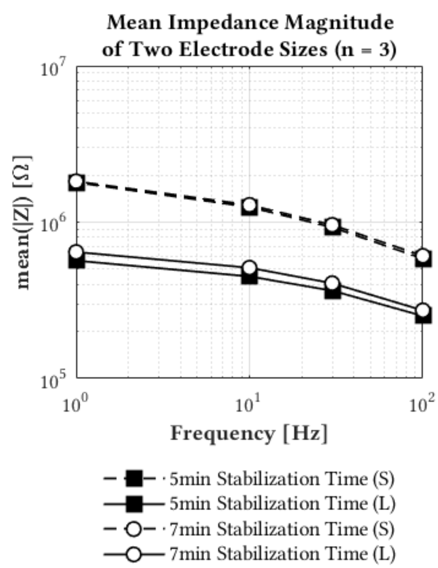


Figure 4. Mean impedance magnitude of the two electrode sizes, small (S) and large (L) at 5 and 7 mins of stabilization times, Stabilization time does not affect impedance magnitude significantly whereas the effect of electrode size on impedance magnitude can be seen clearly

3.2 Electrode Location

In a second experiment, impedances were compared for different positions in the ear canal using a wet skin interface. The mean impedance magnitude from 10 subjects is shown in Figure 5. The mean impedance values for small electrodes at 10 Hz were 28 kΩ (SD = 24 kΩ) for ELE, 42 kΩ (SD = 30 kΩ) for ELJ, and 29 kΩ (SD = 25 kΩ) for ELH respectively whereas for large electrodes the values were 22 kΩ (SD = 21 kΩ) for ELE, 25 kΩ (SD = 14 kΩ) for ELJ, and 29 kΩ (SD = 19 kΩ) for ELH respectively. For visual clarity, the standard deviation was excluded from the plot but can be seen in Table 2 in the supporting material.

L-sized electrodes have a lower impedance in comparison to the S-sized electrodes following the same trend as in dry measurements. The values for wet measurements for all electrodes align with the wet electrode impedance values found in the literature, 34 kΩ (SD = 37 kΩ) [9]. The irregular shape of the ear canal causes different contact pressure on the foam bud and based on anatomy, will likely be the highest at the ELE position [20]. This is also reflected in the impedance results for both sized and contact conditions. Larger electrodes could average out local differences in ear canal surface and the differences between locations becomes smaller. Additionally, the use of ionic solution covering the electrode surface area can reduce the effect of size areas as a liquid film can compensate for skin irregularities. However, fluid might cause issues related to cross-talk between electrodes and hence, there needs to be a tradeoff between size and number of channels. So, the sleep EEG measurements were performed with small size electrode earpieces, ELE as measurement electrode.

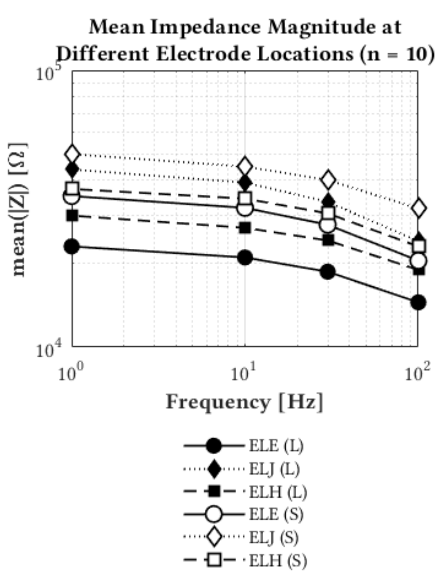


Figure 5. Mean impedance magnitude of the two electrode sizes small (S) and large (L) at locations ELE, ELJ and ELH. Among two sizes, Location ELE shows the lowest impedance compared to the other two locations of both sizes.

3.3 Sleep EEG

The raw signal from the sleep trial was visually investigated using the OpenBCI GUI. Of particular interest were large-signal deviations resulting from subject movements (motion artefacts) as well as signal changes related to the physiological sleep state. Common sleep features include spindles and the slow-wave k-complex. These are seen as bursts in the EEG signal and occur close to one another. The data was filtered with a notch filter at 50 Hz to address line noise as well as a band-pass filter from 1 to 50Hz. It was observed that the Ear-EEG signal was stable and bursts of activity with associated slow waves were present. The signal characteristics are similar to spindle activity from Ear-EEG reported by Mikkelsen et. Al [12]. This provided the first evidence that sleep features could be detected using the Ear-EEG electrode setup.

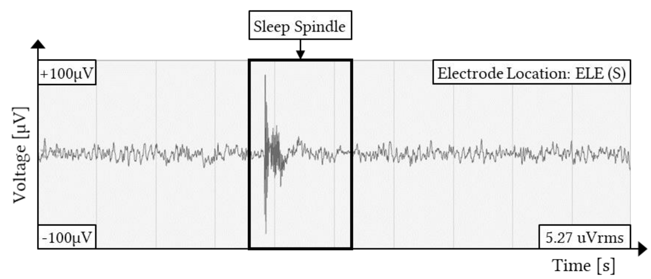


Figure 6. Sleep-EEG signal recorded from the ear during the sleep trial. Bursts in the signal were present and reminiscent of sleep spindle activity as seen in the literature.

4 Conclusion

This study showed the impact of electrode size and location on the skin contact impedance inside the ear-canal for both dry and wet skin contact. Impedances ranged from 450 kΩ to 1.29 MΩ for dry and from 22 kΩ to 42 kΩ for wet contact at 10 Hz and was inversely proportional to the electrode area. The location ELE provided the lowest impedance irrespective of size, likely due to increased attachment pressure to the skin resulting from the anatomy of the ear canal.

We combined small electrode size and the ELE location to detect sleep onset from the ear canal. Characteristic sleep spindles were clearly distinguishable from the baseline during the measurement

The ability to measure sleep activity is ideal for soft electrode designs, as the material can conform will to the ear canal anatomy. During sleep subject movements will be small to minimize motion signal artefacts.

The next step in evaluating soft electrodes includes a comparison to scalp-based EEG to look more closely at correlations between the scalp and ear canal locations and the resulting signal quality.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Moritz Thielen at IDUN Technologies AG for his helpful constructive criticism of the manuscript. We also wish to thank Dr. Katja Junker, Materials Engineer at IDUN Technologies AG for her contribution and help during earpieces production.

REFERENCES

[1] Athavipach, C., Pan-Ngum, S. and Israsena, P. 2019. A wearable in-ear EEG device for emotion monitoring. Sensors (Switzerland). 19, 18 (Sep. 2019), 4014. DOI:https://doi.org/10.3390/s19184014.

[2] Bech Christensen, C., Hietkamp, R.K., Harte, J.M., Lunner, T. and Kidmose, P. 2018. Toward EEG-Assisted Hearing Aids: Objective Threshold Estimation Based on Ear-EEG in Subjects With Sensorineural Hearing Loss. Trends in Hearing. 22, (2018), 1–13. DOI:https://doi.org/10.1177/2331216518816203.

[3] Dong, H., Matthews, P.M. and Guo, Y. 2016. A new soft material based in-the-ear EEG recording technique. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. 2016-Octob, (2016), 5709–5712. DOI:https://doi.org/10.1109/EMBC.2016.7592023.

[4] Goverdovsky, V., Looney, D., Kidmose, P. and Mandic, D.P. 2016. In-ear From Viscoelastic Generic Earpieces: Robust and Unobtrusive 24/7 Monitoring. IEEE Sensors Journal. 16, 1 (2016), 271–277. DOI:https://doi.org/10.1109/JSEN.2015.2471183.

[5] Grozea, C., Voinescu, C.D. and Fazli, S. 2011. Bristle-sensors - Low-cost flexible passive dry EEG electrodes for neurofeedback and BCI applications. Journal of Neural Engineering. 8, 2 (Apr. 2011), 025008. DOI:https://doi.org/10.1088/1741-2560/8/2/025008.

[6] Habibzadeh Tonekabony Shad, E., Molinas, M. and Ytterdal, T. 2020. Impedance and Noise of Passive and Active Dry EEG Electrodes: A Review. IEEE Sensors Journal. 20, 24 (2020), 14565–14577. DOI:https://doi.org/10.1109/JSEN.2020.3012394.

[7] Kalas, M.S. and Momin, B.F. 2016. Stress detection and reduction using EEG signals. International Conference on Electrical, Electronics, and Optimization Techniques, ICEEOT 2016 (Mar. 2016), 471–475.

[8] Kappel, S.L. 2016. Development and Characterization of Ear-EEG for Real-Life Brain-Monitoring. PhD Thesis. 67, 8 (Aug. 2016), 5001–5006. DOI:https://doi.org/10.7146/aul.260.183.

[9] Kappel, S.L. and Kidmose, P. 2015. Study of impedance spectra for dry and wet EarEEG electrodes. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. 2015-Novem, (2015), 3161–3164. DOI:https://doi.org/10.1109/EMBC.2015.7319063.

[10] Kidmose, P., Looney, D., Ungstrup, M., Rank, M.L. and Mandic, D.P. 2013. A study of evoked potentials from ear-EEG. IEEE Transactions on Biomedical Engineering. 60, 10 (2013), 2824–2830. DOI:https://doi.org/10.1109/TBME.2013.2264956.

[11] Matthies, D.J.C. 2013. InEar BioFeedController. CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13 (New York, New York, USA, 2013), 1293.

[12] Mikkelsen, K.B., Kappel, S.L., Hemmsen, M.C., Rank, M.L. and Kidmose, P. 2019. Discrimination of Sleep Spindles in Ear-EEG. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS (Jul. 2019), 6697–6700.

[13] Mikkelsen, K.B., Tabar, Y.R., Kappel, S.L., Christensen, C.B., Toft, H.O., Hemmsen, M.C., Rank, M.L., Otto, M. and Kidmose, P. 2019. Accurate whole-night sleep monitoring with dry-contact ear-EEG. Scientific Reports. 9, 1 (2019), 1–12. DOI:https://doi.org/10.1038/s41598-019-53115-3.

[14] Nakamura, T., Goverdovsky, V., Morrell, M.J. and Mandic, D.P. 2017. Automatic Sleep Monitoring Using Ear-EEG. IEEE Journal of Translational Engineering in Health and Medicine. 5, July (2017), 1–8. DOI:https://doi.org/10.1109/JTEHM.2017.2702558.

[15] Oh, S.Y., Leung, L., Bommannan, D., Guy, R.H. and Potts, R.O. 1993. Effect of current, ionic strength and temperature on the electrical properties of skin. Journal of Controlled Release. 27, 2 (1993), 115–125. DOI:https://doi.org/10.1016/0168-3659(93)90215-Q.

[16] PalmSens Product: 2021. https://www.palmsens.com/product/palmsens4/. Accessed: 2021-06-21.

[17] PalmSens Software: 2021. https://www.palmsens.com/software/ps-trace/. Accessed: 2021-06-22.

[18] Paul, A., Deiss, S.R., Tourtelotte, D., Kleffner, M., Zhang, T. and Cauwenberghs, G. 2019. Electrode-Skin Impedance Characterization of In-Ear Electrophysiology Accounting for Cerumen and Electrodermal Response. International IEEE/EMBS Conference on Neural Engineering, NER. 2019-March, (2019), 855–858. DOI:https://doi.org/10.1109/NER.2019.8716918.

[19] Puurtinen, M.M., Komulainen, S.M., Kauppinen, P.K., Malmivuo, J.A.V. and Hyttinen, J.A.K. 2006. Measurement of noise and impedance of dry and wet textile electrodes, and textile electrodes with hydrogel. Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings. August (2006), 6012–6015. DOI:https://doi.org/10.1109/IEMBS.2006.260155.

[20] Valentin, O., Viallet, G., Delnavaz, A., Cretot-Richert, G., Ducharme, M., Monsarat-Chanon, H. and Voix, J. 2021. Custom-fitted in-and around-the-ear sensors for unobtrusive and on-the-go eeg acquisitions: Development and validation. Sensors. 21, 9 (Apr. 2021), 2953. DOI:https://doi.org/10.3390/s21092953.

[21] Zhou, M., Tian, C., Cao, R., Wang, B., Niu, Y., Hu, T., Guo, H. and Xiang, J. 2018. Epileptic Seizure Detection Based on EEG Signals and CNN. Frontiers in Neuroinformatics. 12, December (Dec. 2018), 1–14. DOI:https://doi.org/10.3389/fninf.2018.00095.

Coremoni-WE: Individual Core Training Monitoring and Support System Using an IMU at the Waist and the Ear

Nishiki Motokawa
Aoyama Gakuin University
Tokyo, Japan
nmotokawa@wil-aoyama.jp

Ami Jinno
Aoyama Gakuin University
Tokyo, Japan
ajinno@wil-aoyama.jp

Yushi Takayama
Aoyama Gakuin University
Tokyo, Japan
ytakayama@wil-aoyama.jp

Shun Ishii
Aoyama Gakuin University
Tokyo, Japan
sishii@wil-aoyama.jp

Anna Yokokubo
Aoyama Gakuin University
Tokyo, Japan
yokokubo@it.aoyama.ac.jp

Guillaume Lopez
Aoyama Gakuin University
Tokyo, Japan
guillaume@it.aoyama.ac.jp

ABSTRACT

Due to the influence of the new coronavirus, many people are interrupting fitness clubs and exercise/sports performed by multiple people. Under these circumstances, "core training," which can be easily performed indoors by individuals, attracts attention as an exercise to improve health. However, it is not easy to recognize whether the posture during the training is correct or not, which may significantly reduce the effect of the exercise. To tackle these issues, we have been developing "Coremoni-WE," a core training monitoring and support system that combines wearables and earables to judge posture.

CCS CONCEPTS

• **Computer systems organization → Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

wearable device, sports, fitness monitoring

ACM Reference Format:

Nishiki Motokawa, Ami Jinno, Yushi Takayama, Shun Ishii, Anna Yokokubo, and Guillaume Lopez. 2021. Coremoni-WE: Individual Core Training Monitoring and Support System Using an IMU at the Waist and the Ear. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3460418.3479325

1 INTRODUCTION

In recent years, people have been spending more and more time at home to prevent the spread of new coronaviruses. In a survey on health[2], more than 30% of the respondents answered that they became more active after the coronavirus outbreak, and "core training," which can be easily performed indoors by individuals, is attracting attention. However, there is a concern that the effects of trunk training performed by individuals may be reduced because the trunk is often not used correctly compared to the activity conducted under an instructor.This study proposes a support system for core training using wearable and earable IMUs to help individuals perform core training correctly.

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

ACM ISBN 978-1-4503-8461-2/21/09…$15.00
https://doi.org/10.1145/3460418.3479325

2 RELATED WORKS

Barbado et al. studied about the reliability of smartphone accelerometers to quantify the intensity of core training [1]. From the five types of core training performed by 23 participants, they obtained moderate-to-high reliability scores for pelvic acceleration and concluded that smartphone accelerometers is useful to identify the individuals' core training status and to improve the training programs. However, this study is not mentioned about the correct postures and user feedbacks.

While many researchers and developers have been developing applications based on smartphones and smartwatches, Kawsar et al. [3] proposed and developed a new wearable platform called "eSense". The eSense platform consists of a pair of wireless earbuds augmented with kinetic, audio, and proximity sensing [4]. The left earbud has a six-axis IMU with an accelerometer, a gyroscope, and a Bluetooth Low Energy (BLE) interface to stream sensor data to a paired smartphone. Both earbuds are also equipped with microphones to record external sounds.

Prakash et al. introduced the advantages of eSense in counting the number of steps of walking [5]. While the head movement can still pollute this bouncing signal, they proposed a method to alleviate the noise from head movement. Results show 95% step count accuracy even in the most challenging test case—very slow walk—where smartphone and wrist-band-type systems falter. Importantly, their system, which is named "STEAR" (STep counting from EARables) is robust to changes in walking patterns and scales well across different users. Additionally, they demonstrated how STEAR also brings opportunities for practical jump analysis, often crucial for exercises and injury-related rehabilitation.

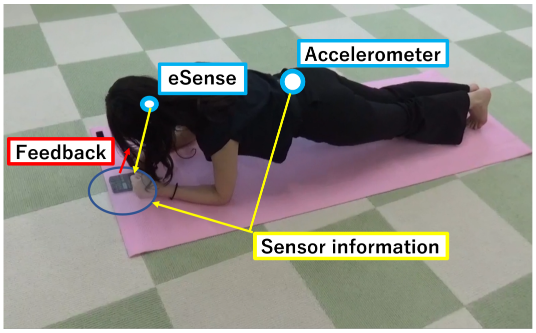
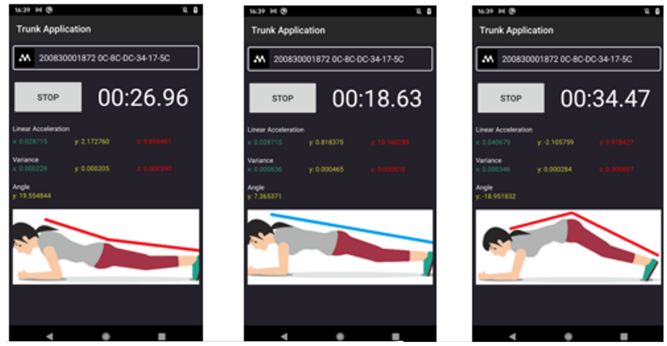


Figure 1: configuration of CoreMoni-WE



When the waist is down (angle less than 5 degrees)

When the posture is correct (angle of more than 5 degrees but less than 15 degrees)

When the hips are up (angle over 15 degrees)

Figure 2: CoreMoni-WE's visual feedback for front plank

3 PROPOSED SYSTEM

3.1 CoreMoni Overview

CoreMoni is a core training monitoring and support system that focuses on the user's "posture" and "trunk blurring" in multiple core training exercises. It guides to improve the "posture" and "trunk blurring." In this study, we selected a method to determine "appropriate posture" using wearable and earable sensors (see usage image in Fig. 1). CoreMoni collects the acceleration and angular velocity values of the trunk and head from a wearable acceleration sensor attached to the user's waist (movesense[1]) and an earable acceleration sensor attached to the head (eSense[2]), and transmits the data to a mobile information terminal (smartphone or tablet PC). The transmitted data is converted to variance and angle values and used for feedback on the mobile terminal. Two types of feedback are provided: switching images for "posture" and sound effects for "blur of the trunk" during the trunk training.

3.2 Examination of the number of accelerometers installed and their positions

We conducted a requirement extraction experiment to determine the number and position of accelerometers in the CoreMoni-WE implementation. One university student in his 20's was asked to perform front plank, which is the most basic core training exercise. We attached accelerometers to the subject's head, waist, and ankles and asked him to perform front plank for one minute according to the following procedure.

- 0 - 15 sec: correct posture
- 15 - 30 sec: low back posture
- 30 - 45 sec: high back posture
- 45 - 60 sec: continuously changing the posture

We recorded the acceleration at each sensor position during the elapsed time of 1 minute of core training. After the experiment, the results were plotted for analysis. Acceleration of the sensors attached to the head and the ankle does not change much along any axis. On the other hand, the acceleration of the sensor attached to the waist showed some change, but the difference was significant only in the y-axis. We can assume that the y-axis of acceleration at the waist is usefull to detect "trunk blur." Since two points are preferable to estimate trunk posture, and no difference was observed between acceleration at the ankle and the head, in addition to the wearable sensor at the waist, we selected the earable sensor at the head, assuming it is more likely to be used than a wearable sensor at the ankle. A feedback application based only on the waist attached wearable sensor output as been developed, using thresholds on y-axis acceleration for evaluating "posture" (see Fig. 2).

4 FUTURE WORKS

This paper proposed supporting individual core training with a system using accelerometers attached to the waist and ear. Our first application prototype can judge front plank posture using only waist attached wearable sensor. We are designing new algorithms combining waist attached wearable sensor and head attached earable sensor to discriminate and support multiple core training exercises. In this study, we focused only on acceleration among the data that can be collected from wearable and earable sensors. Still, in the future, it may be possible to perform training more following the user's health condition by collecting biological signals such as heart rate.

ACKNOWLEDGMENTS

This work was supported by Aoyama Gakuin University Research Institute grant program for creation of innovative research.

REFERENCES

[1] David Barbado, Belen Irles-Vidal, Amaya Prat-Luri, María Pilar García-Vaquero, and Francisco J. Vera-Garcia. 2018. Training intensity quantification of core stability exercises based on a smartphone accelerometer. *PLOS ONE* 13, 12 (12 2018), 1–13. https://doi.org/10.1371/journal.pone.0208262

[2] Meiji Yasuda Life Insurance. 2020. Survey on "Health" - Change in health awareness and behavior after the Stay Home Corona disaster! . https://www.meijiyasuda.co.jp/profile/news/release/2020/pdf/20200902_01.pdf

[3] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Allesandro Montanari. 2018. Earables for personal-scale behavior analytics. *IEEE Pervasive Computing* 17, 3 (2018), 83–89.

[4] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Alessandro Montanari. 2018. Earables for Personal-Scale Behavior Analytics. *IEEE Pervasive Computing* 17, 3 (2018), 83–89. https://doi.org/10.1109/MPRV.2018.03367740

[5] Jay Prakash, Zhijian Yang, Yu-Lin Wei, and Romit Roy Choudhury. 2019. STEAR: Robust Step Counting from Earables. In *Proceedings of the 1st International Workshop on Earable Computing* (London, United Kingdom) *(EarComp'19)*. Association for Computing Machinery, New York, NY, USA, 36–41. https://doi.org/10.1145/3345615.3361133

[1] movesense, SUUNTO, https://www.movesense.com/

[2] eSense, Nokia Bell Labs, Cambridge, https://www.esense.io/

PilotEar: Enabling In-ear Inertial Navigation

Ashwin Ahuja
aa2001@cl.cam.ac.uk
University of Cambridge
United Kingdom

Andrea Ferlini
af679@cl.cam.ac.uk
University of Cambridge
United Kingdom

Cecilia Mascolo
cm542@cl.cam.ac.uk
University of Cambridge
United Kingdom

ABSTRACT

Navigation systems are used daily. While different types of navigation systems exist, inertial navigation systems (INS) have favorable properties for some wearables which, for battery and form factors may not be able to use GPS. Earables (*aka* ear-worn wearables) are living a momentum both as leisure devices, and sensing and computing platforms. The inherent high signal to noise ratio (SNR) of ear-collected inertial data, due to the vibration dumping of the musculoskeletal system; combined with the fact that people typically wear a pair of earables (one per ear) could offer significant accuracy when tracking head movements, leading to potential improvements for inertial navigation. Hence, in this work, we investigate and propose PilotEar, the first end-to-end earable-based inertial navigation system, achieving an average tracking drift of $0.15\frac{m}{s}$ for one earable and $0.11\frac{m}{s}$ for two earables.

CCS CONCEPTS

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

KEYWORDS

earables; inertial navigation; calibration; dataset; sensor fusion

ACM Reference Format:
Ashwin Ahuja, Andrea Ferlini, and Cecilia Mascolo. 2021. PilotEar: Enabling In-ear Inertial Navigation. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3460418.3479326

1 INTRODUCTION

Navigation systems are ubiquitous, they are in our phones, cars, sometimes even in our smartwatches. People rely on navigation systems daily: commuting, while running errands, when going to meet friends and family, driving to a restaurant, etc. Broadly speaking, it is possible to classify navigation systems as satellite-based [36], or inertial based [32]. Satellite-based navigation systems are aided by, for example, a Global Positioning System (GPS). On the other hand, Inertial Navigation Systems (INS) leverage inertial measurement units (IMUs) to maintain the location of a device without the need for any satellite device. While GPS-based navigation systems generally have greater accuracy than INS, they have shortcomings when it comes to battery life, and sometimes (e.g., indoors where GPS coverage is limited) fail to obtain the GPS lock they require to function. Unlike satellite-based navigation systems, INS are well suited to applications where either GPS coverage may be limited or where battery life is a concern.

In this work, for the first time, we explore the potential of an *earable*-based inertial navigation system. Earables (also know as ear-worn wearables) is an exploding area for wearable research. They allow users to combine listening to music with sensing and computing. Specifically to this work, earables offer significant potential benefits for inertial tracking, a key part of an inertial navigation system. Concretely, earables can be used to track head movements which, in turn, can act as a proxy for visual attention [9]. As a result of that, earable could be effectively used as an navigation system by people visually impaired to navigate through audio feedback. Additionally, the stabilization effect of the human musculoskeletal postural system ensures a natural vibration damping [14] which leads to reduced noise in the motion sensor data fed to the INS. Furthermore, unlike other wearables, earables present the unique opportunity of two distinct vantage points (one from each ear), which can be leveraged to increase the overall performance of the system.

However, to implement an earable-based navigation system, a number of challenges needs to be overcome. First and foremost the lack of an existing earable platform equipped with a magnetometer [8]: we prototype a new ear-worn sensing platform which leverages a powerful an Arduino Nano 33 BLE Sense as microcontroller (MCU). Our prototype earable (Figure 1) collects motion data from the 9-axis IMU on-board the MCU, and streams it over Bluetooth Low Energy (BLE). Further, it provides audio feedback thanks to a custom-made printed circuit board (PCB) which produces audible tones using pulse width modulation (PWM) and a resistor. Our prototype is designed to use minimal power and can run for over a day of continuous use. It also contains other sensors (temperature, pressure, microphone) which we do not experiment with in this paper. Secondly, because of the novelty of earable computing, there is a dataset scarcity. We therefore run a small user study and collect IMU data using our prototype. Thirdly, it is well known that inertial approaches require calibration in order to provide accurate results [12]. Thus, we build upon previous work to define an effective calibration framework to calibrate the 9 IMU axis of our prototype. Specifically, we leverage the work done by Sipos et al. [29] to calibrate the accelerometer, we implement the gyroscope calibration by Shen et al. [28], and finally we calibrate the magnetometer following the approach presented by Ferlini et al. [8]. Lastly, existing position tracking techniques have to be adapted to earables. We find that a heading estimation approach which leverages the combination of a gyroscopic heading and a

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

ACM ISBN 978-1-4503-8461-2/21/09.
https://doi.org/10.1145/3460418.3479326

Figure 1: Earable prototype

magnetometer heading fused together with a complementary filter outperforms existing fusion algorithms. Correct heading estimation which successfully tracks the position of a person in a navigation application is predicated on displacement estimation. We perform this estimation by means of a pedestrian dead reckoning algorithm adapted over Lu et al. [17].

The contributions of this work can be summarized as follows.

- We prototype an earable platform for collecting and transmitting accelerometer, gyroscope, and magnetometer data over Bluetooth Low Energy (BLE) and audio feedback. We use such prototype to collect a new in-the-wild dataset consisting of 9-axis IMU data both from the earable prototype and an iPhone. The user study has been carried out in agreement with the university ethics committee, and comprises data of six users walking both indoors and outdoors, both in low noise and high noise situations. To the best of our knowledge, there is not a similar publicly available dataset, and we will share ours the research community.
- We implement and evaluate the performance of PilotEar, the first end-to-end inertial navigation system for earables, achieving an average drift of $0.15\frac{m}{s}$ for one earable and $0.11\frac{m}{s}$ when fusing both earables.
- We experiment with a novel multi-device sensor fusion approach to combine both earables and a smartphone. The proposed approach reduces power usage for accurate tracking by 65% by incorporating occasional GPS updates, whilst reducing tracking drift by 27% when maximizing performance.

2 INERTIAL NAVIGATION PRIMER

Inertial Navigation generally consists of three stages.

The first stage is the **pre-processing** stage. IMU data are sampled and then pre-processed to remove the noise and the temporal drift which often affect IMU readings. The first challenge is to effectively calibrate each sensor. In this work, the sensors we consider, and calibrate, are accelerometer, gyroscope, and magnetometer.

Once the IMU data are successfully pre-processed and a calibration framework has been applied, Secondly, the pre-processed data is used to find the heading of the individual. This second phase goes by the name **heading estimation** and, in the case of earables consists in estimating the orientation of the devices. The heading can be estimated either from the gyroscope [31], from the magnetometer [8], or fusing the both. There are many ways to fuse gyroscope and magnetometer data. In this work we fuse gyroscopic and magnetic data in a similar way to what Shen et al. [28] did by mean of a complementary filter. Specifically, the gyroscopic and magnetic heading are estimated independently and then are fused as follows:

$$\theta = (\mathbf{1} - \alpha)\theta_\omega + \alpha\theta_\mathbf{m} \tag{1}$$

Where θ_ω and θ_m are the gyroscopic and magnetic heading, respectively. The choice of α determines how much of each method is used. Shen et al. [28] suggest using a low value to fuse the gyroscope and magnetometer heading. Other papers suggest increasing the value of α over time, as the gyroscopic drift increases. Although not featured in this paper, other well known fusion algorithms are the Madgwick filter [19], the Mahony filter [20], and the Fourati filter [10]. We note that our preliminary results suggest the complementary filter approach adapted from Shen et al. [28] outperforms our implementation of all the Madgwick, the Mahony and the Fourati filters.

Finally, the third and last phase consists in finding the change in position (i.e., displacement) of the user at any given timestamp. This phase is also known as **displacement estimation** phase. Concretely, the position of the user is estimated by leveraging the their heading and linear acceleration to find the change in their (x, y) position at each timestamp. We do that by adapting an existing pedestrian dead reckoning model to complete the step length estimation effectively on earables. In pedestrian dead reckoning (PDR), the goal is to identify every step that the user makes when walking. This is used in practice to identify the distance by which the user has moved at each time step, given the time taken to walk a known length stride. Our adaptation builds upon the step length estimation proposed by Lu et al. [17]:

$$\mathbf{D} = K \cdot \iint \boldsymbol{a}(t)dtdt \tag{2}$$

This assumes the velocity for each step starts at 0, whilst the value of K is computed experimentally for each user.

3 SYSTEM DESIGN

The lack of an existing programmable earable platform equipped with a magnetometer compelled us to build to a new earable prototype platform. Our prototype features a 9-axis IMU, alongside a microphone, a temperature sensor, and pressure sensor. The device is easily re-programmable using the Arduino prototyping platform. We wanted the prototype to be self-contained, thus we included a battery and power management. To be useful in real-world applications, the device should also be able to output audio.

We prototyped our earable platform around an Arduino BLE Sense 33. This, in turn, runs on a powerful ARM Cortex M4 microcontroller, and is equipped with a BLE transceiver, 9-axis IMU, microphone, temperature and pressure sensors. For battery and power management we used a 520mAh lithium polymer battery (LiPo) and power regulation module (PAM2401 as a DC-DC boost converter and MCP73831 as a charging regulator). To enable the headphone management and switching circuitry (using a TS3A24159), not natively supported by the Arduino BLE Sense 33, we designed and printed a two-layer custom PCB (Figure 2). The PCB design (created using EagleCAD) will be released so that researchers can alter the board for their requirements if necessary.

By using our prototype, researchers can easily collect and transmit data via Bluetooth to a connected device. The ubiquitous Arduino platform can be used to write software for this, exploiting the significant existing open-source code. We found that we could transmit 9-axis IMU from the prototype earable to a connected an Apple iPhone 11, at 40Hz, with a battery life of over one day (26.3 hours), and a current draw of 20.9mA.

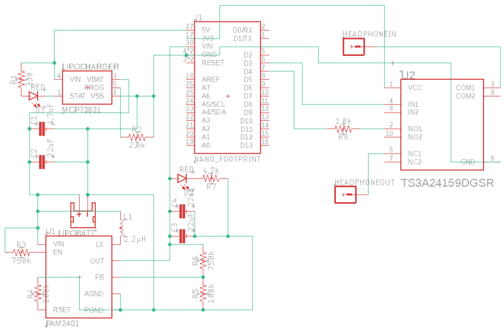


Figure 2: Prototype schematics

As opposed to other earable platforms, like eSense [15], which are designed as in-ear headphones using injection moulded plastic, our prototype was 3D-printed, with an around-ear design. With the current printing material of polylactic acid (PLA), the sizing is user specific (Figure 4). However, if the design could be printed in a more flexible material such as thermoplastic polyurethane (TPU), the design could be more universal. Figure 3 shows the prototype design while Figure 1 shows the complete system which we designed and used to collect data.

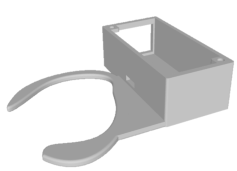


Figure 3: Case design.

Figure 4: Prototype worn.

4 EARABLE NAVIGATION

PilotEar consists of a number of elements. First, sensor calibration. The aim of this phase is to remove biases and temporal drift from the raw IMU data. Following this first step, there is heading estimation, where the calibrate IMU data are used to estimate the direction where the user is facing. Since we were only concerned with motion on the plane of the earth's surface, we did not need to deal with any other orientation axis. The third aspect aims at estimating the distance moved per time step. These three phases, once combined, enable user position tracking from single earable. To improve the performance of PilotEar, we also considered several methods to fuse the data sampled by both earables. Finally, we present a method to provide audio feedback, for instance to a visually impaired user. By doing so, we believe PilotEar could help in guiding a visually impaired person towards the correct exit at a complicated intersection.

4.1 Sensor Calibration

We calibrated the accelerometer using Sipos et al.'s calibration model (Equation (3)) [29]. We minimized the square error for calibration clips with each earable in a number of static orientations using the Levenberg-Marquandt Algorithm.

$$\mathbf{a}' = \begin{pmatrix} 1 & 0 & 0 \\ \alpha_{yx} & 1 & 0 \\ \alpha_{zx} & \alpha_{zy} & 1 \end{pmatrix} \begin{pmatrix} \mathrm{SF}_{ax} & 0 & 0 \\ 0 & \mathrm{SF}_{ay} & 0 \\ 0 & 0 & \mathrm{SF}_{az} \end{pmatrix} \times \left(\mathbf{a} - \begin{pmatrix} b_{ax} \\ b_{ay} \\ b_{az} \end{pmatrix} \right) \tag{3}$$

To calibrate the gyroscope, we used a method described by Shen et al. [28]. Here, when the device is relatively stationary ($|\mathbf{acc}| \approx g$), the gyroscope is calibrated using the magnetometer. To do so, we integrate the gyroscopic headings over time to find the orientation of the device. We then find a rotational offset between the gyroscopic heading and the magnetometer heading.

Finally, for magnetometer calibration, we investigate the real-world tracking efficacy of Ferlini et al.'s semi-automated calibration, which uses occasional phone reference points to calibrate an earable offset [8]. This is based on the idea that when a person unlocks and uses their smartphone, the phone and earable are aligned. Hence, the reference heading from the phone (trustworthy) can be used to correct the magnetometer of the earable. We emulate this by collecting potential calibration points only every 15s. Through preliminary testing, we define a regime to maintain a rolling window of up-to fifteen calibration points, but completing a full re-calibration if we have encountered a full rotation (when the heading rolls over from 359° to 0° or vice versa).

4.2 Heading Estimation

Our results suggests the best heading estimation comes from simply using the calibrated magnetometer:

$$\psi = \arctan \frac{m_y}{m_x} \tag{4}$$

To improve the accuracy, we added tilt compensation, to account for the fact that the earth's magnetic field does not run perpendicularly to the surface other than at the equator. We do this by finding pitch and roll (ϕ and θ, respectively) from the accelerometer, when stationary.

$$\phi = \arctan \frac{a_y}{a_z} \tag{5}$$

$$\theta = \arctan \frac{-a_x}{a_y \sin\phi + a_z \cos\phi} \tag{6}$$

θ and ϕ are then used to rotate the magnetometer readings to the flat plane, where $\theta = \phi = 0$

Whilst offering higher accuracy, the magnetometer-only approach suffers from a potentially lower reliability when a calibration fails. We, therefore, combined the magnetometer-only heading with the gyroscope-only heading by mean of a complementary filter. α is set to $0.8 - \frac{t}{400}$.

$$\psi = (0.8 - \frac{t}{400}) \cdot \text{heading\_gyro} + (0.2 + \frac{t}{400}) \cdot \text{heading\_mag} \tag{7}$$

Concretely, we trust the gyroscope more at the beginning. Over time, the magnetometer calibration is likely to be more refined and, at the same time, the gyroscopic bias may arise. Hence the increased trust on the magnetometer.

4.3 Displacement Estimation

For the displacement estimation, we adapted Lu et al.'s PDR approach [17] to identify the stride. We use a constant stride length defined by the height of the user:

$$\text{Stride\_Length} = 0.43 \cdot \text{height} \tag{8}$$

To find each stride, we used the following algorithm:

- Low-pass filter accelerometer norm with 3Hz cut-off frequency.
- Find peaks (local maxima) of the filtered data.
- Find peaks with topographic prominence above a threshold.

Given the time of each stride, we find the distance moved per timestamp and the change in position. For each stride occurring between timestamp i and j:

$$d_{t \in i,\ldots,j} = \frac{\text{Step\_Length}}{j - i} \tag{9}$$

Then, for each timestamp:

$$\mathbf{S}_t = \begin{pmatrix} S_{t_x} \\ S_{t_y} \end{pmatrix} = \mathbf{S}_{t-1} + \begin{pmatrix} d_t \cos \psi_t \\ d_t \sin \psi_t \end{pmatrix} \tag{10}$$

4.4 Sensor Fusion

In combining the data from the two earables, we consider methods to improve the accuracy and reduce the power consumption of the tracking.

To improve the accuracy, we combine the headings of the two earables with a particle filter. The complete algorithm for accurate tracking using two earables can be summarized as:

- Calibrate both earables independently.
- Find the heading at each timestamp for each earable using a complementary filter of gyroscopic and magnetometer heading.
- Combine the headings with a particle filter to identify the most probable heading at each timestamp.
- Find the stride time according to each earable with Algorithm 1. Average these for both earables.
- Find the position according to Equation 10.

To reduce the power consumption, we propose two methods. First, we consider dropping the sampling frequency of one earable (to 5Hz/ 10Hz), whilst maintaining the frequency of the other at 20Hz. We demonstrate that the battery life would increase by 33 minutes when dropping the data frequency from $20Hz$ to $10Hz$, increasing by another 17 minutes going to $5Hz$. In the second method, we use GPS updates every 30 seconds. We take the error model of GPS to be a Gaussian distribution with $\sigma = 3.9m$ since the US claim a 95% accuracy of $7.8m$ [26]. This is used as the re-sampling probability for a positional particle filter. This allows us to maintain a maximum average error of $15m$. Therefore, by using the current figures from a NEO-6 GPS module, the ability to return the GPS to idle (when a lock is maintained) before using it at high performance for $1s$ would reduce the power usage of a smartphone positional tracking system by 65%.

4.5 Audio Feedback

To demonstrate a potential use-case for tracking using earables, we created an iPhone application to help guide users to a target heading. The difference between the calibrated heading and the target heading was found and converted into a tone as follows:

$$f = 2750\left(\frac{180 - \text{heading\_diff}}{180}\right) + 250 \tag{11}$$

By turning their head, the user could audibly gain an understanding of the correct direction, with the tone getting higher in pitch as they are facing the right direction.

5 DATA COLLECTION

We recruited six healthy subjects (balanced in gender). The participants wore two earable prototypes, one per ear, and held an iPhone in front of them. The system logged 9-axis IMU data and reference heading data from the iPhone. We collected data while the users walked two different circuits. The first was indoors, walking up and down a $10m$ corridor five times. The second involved walking outside, where the levels and composition of noise would differ, around a $750m$ route. In both cases, the test was run twice. The experiment was granted permission by our university's ethics committee.

Figure 5: Subject completing the indoor portion of user study (left) and outdoor portion (right).

6 EVALUATION

We test our system on two different metrics. The first is the difference between the predicted heading and the ground truth reference heading. These are averaged across all the subjects. We also look at the drift in tracking position over time (in $\frac{m}{s}$). Since GPS could not be used to provide ground truth for the indoor test, both the indoor and outdoor tests were designed to start and end at the same point. Hence, the drift is the normal of the final predicted position divided by the testing time.

6.1 Single Earable Tracking

We first evaluate our calibration framework, showing that the accelerometer calibration, magnetometer calibration and gyroscope calibration methods performed statistically significantly better (using a paired t test with $\alpha = 0.05$) than having no calibration, irrespective of the method used for heading estimation. In particular, calibrating the magnetometer leads to a 38% improvement over no calibration. We also demonstrate the efficacy of our complementary filter-based heading estimation method, showing how it achieves amongst the lowest heading and displacement errors with lowest

Table 1: Results comparing heading error for different heading estimation methods.

	Heading Error / °		Displacement Error / ms^{-1}	
	Average	Standard Deviation	Average	Standard Deviation
Magnetometer only	15.1	15.7	0.133	0.0354
Gyroscope only	21.9	15.7	0.194	0.0217
Madgwick Filter	19.5	15.7	0.228	0.0423
Fourati Filter	21.9	16	0.181	0.0275
Complementary Filter	15.8	14.0	0.132	0.0214

Table 2: Results comparing displacement error for different displacement estimation methods.

	Heading Error / °		Kinematics Displacement Error / ms^{-1}		Step Length Estimation Displacement Error / ms^{-1}	
	Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation
Magnetometer only	15.1	15.6	2.32	0.769	0.133	0.0354
Gyroscope only	21.9	15.7	2.69	1.11	0.194	0.0217
Madgwick Filter	19.4	15.7	3.66	1.55	0.228	0.0423
Fourati Filter	21.9	16	2.42	1.14	0.181	0.0275
Complementary Filter	15.8	14.0	2.63	0.538	0.132	0.0214

standard deviations (Table 1). This denotes the intuition of weighting the accuracy of magnetometer methods, with better reliability worked well. In particular, it statistically significantly outperforms existing sensor fusion algorithms such as Madgwick and Fourati filters. Further, we compare our displacement estimation method against a simple kinematics approach where the velocity is maintained over time, not exhibiting a velocity drift that was shown by the kinematics method. Table 2 shows how the proposed method outperforms the kinematics one.

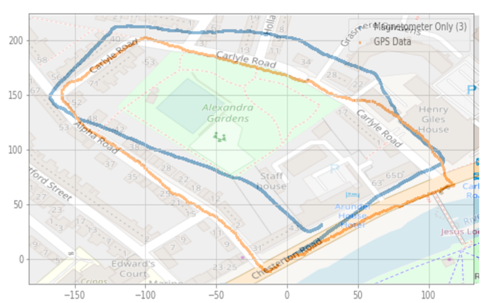


Figure 6: Example of outdoor tracking path.

Figure 6 reports an example path found by tracking a single earable on the outdoor test, showing how our method for single earable tracking performed very well, with a minimum average tracking drift of around 0.15$\frac{m}{s}$.

6.2 Two Earable Tracking

Table 3 demonstrates how combining the data from both earables statistically significantly outperforms using one earable. This difference is more notable outside, likely due to the additional noise. Our results (Table 4 and Figure 7) suggest that dropping the data frequency of one earable generally adversely impacted the displacement error. For indoor tests, there was a statistically significant difference going from 20Hz to 10Hz and 10Hz to 5Hz, however,

Table 3: Results for combining both earables when particle filtering the heading.

	Indoors Final Displacement Error / ms^{-1}		Outdoors Final Displacement Error / ms^{-1}	
	Average	Standard Deviation	Average	Standard Deviation
Left Earable	0.173	0.096	0.181	0.025
Right Earable	0.135	0.081	0.170	0.082
Reference Heading	0.077	0.038	0.104	0.061
Both Earables	0.119	0.089	0.021	0.039

Table 4: Results for mixing timings for earables.

	Indoors Final Displacement Error / ms^{-1}		Outdoors Final Displacement Error / ms^{-1}	
	Average	Standard Deviation	Average	Standard Deviation
Left Earable	0.173	0.096	0.226	0.025
Right Earable	0.135	0.081	0.139	0.082
Reference Heading	0.077	0.038	0.168	0.061
Both 20Hz	0.119	0.089	0.106	0.039
One 20Hz, one 10Hz	0.128	0.094	0.227	0.102
One 20Hz, one 5Hz	0.156	0.081	0.204	0.049
One 20Hz, one 2.5Hz	0.185	0.119	0.242	0.119

there was no difference going from 5Hz to 2.5Hz. Additionally, the results for 10Hz and 5Hz still performed better than one earable individually. For the outdoor tests, the changes in data frequency had a less obvious effect. There was no clear trend with the reducing the second earable's frequency. However, for all frequency reductions, there was a significant negative impact. We suspect that this arose from an increased background noise that came from the outdoor environment. This noise meant that additional measurements were important for tracking accuracy.

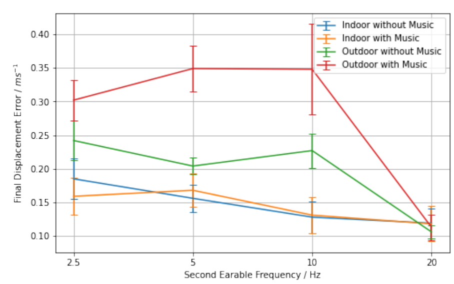


Figure 7: Tracking drift when mixing timings of earables.

7 RELATED WORK

Earable research is living a momentum [6]. In the literature earables have been used for a number of applications ranging form head motion tracking [8, 9], to inertial sensing for activity recognition [15] and step counting [24], acoustic sensing [18], and a number of mobile health applications. A few (non comprehensive list of) examples are dietary habits monitoring [1–3], teeth grinding (Bruxism)/jaw clenching [27] detection, body-core temperature and blood pressure monitoring [4], in-ear photoplethysmography (PPG) [22, 23], and sleep monitoring [21]. In this work, for the first time, we leverage earables to build an end-to-end navigation system.

There is a significant body of related work for inertial navigation. This includes papers by Shen et al. [28], Woodman et al. [34], and

Kok et al. [16] who create full INS using 9-axis IMUs. There are also significant works which focus on specific components of the process, some for wearable devices. Won et al. [33], Frosio et al. [11] and Skog et al. [30] look at accelerometer calibration. For gyroscope calibration, there is significant work on calibration methods that require precise turntables, including work by Yang et al. [35] and Chen et al. [5]. Finally, for magnetometer calibration, there is work by Renaudin et al. [25]. We instead use the method defined for earables by Ferlini et al. [8]. For heading estimation, there are numerous approaches, some of which are considered in this paper. These include methods by Madgwick et al. [19] and Fourati et al. [10]. Other prominently used methods include Mahony et al. [20] and QUEST [7]. For displacement estimation, there is previous work on methods which work well for wearables, using features of hand movement [28], and foot movement [13] to improve performance. However, to the best of our knowledge, there is no existing work which carries out any inertial navigation on earables.

8 FINAL REMARKS

In this work we introduce PilotEar, a novel framework for earable-based inertial navigation. PilotEar is capable of achieving a drift of as little as $0.11\frac{m}{s}$, when fusing the data coming from both ears. Our work defines an effective calibration schema, as well as precise heading and displacement estimation methods. Ultimately, we believe that PilotEar, providing acoustic feedback, could offer significant benefits for navigation, particularly for visually impaired people.

ACKNOWLEDGMENTS

This work is supported by Nokia Bell Labs through their donation for the Centre of Mobile, Wearable Systems and Augmented Intelligence.

REFERENCES

[1] Oliver Amft, Mathias Stäger, Paul Lukowicz, and Gerhard Tröster. 2005. Analysis of chewing sounds for dietary monitoring. In *International Conference on Ubiquitous Computing*. Springer, 56–72.

[2] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. 2017. EarBit: using wearable sensors to detect eating episodes in unconstrained environments. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1, 3 (2017), 1–20.

[3] Shengjie Bi, Tao Wang, Nicole Tobias, Josephine Nordrum, Shang Wang, George Halvorsen, Sougata Sen, Ronald Peterson, Kofi Odame, Kelly Caine, Ryan Halter, Jacob Sorber, and David Kotz. 2018. Auracle: Detecting Eating Episodes with an Ear-mounted Sensor. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–27. https://doi.org/10.1145/3264902

[4] Nam Bui, Nhat Pham, Jessica Jacqueline Barnitz, Zhanan Zou, Phuc Nguyen, Hoang Truong, Taeho Kim, Nicholas Farrow, Anh Nguyen, Jianliang Xiao, et al. 2019. eBP: A wearable system for frequent and comfortable blood pressure monitoring from user's ear. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–17.

[5] Yi Chen, Ethem Erkan Aktakka, Jong-Kwan Woo, Khalil Najafi, and Kenn Richard Oldham. 2018. On-chip capacitive sensing and tilting motion estimation of a micro-stage for in situ MEMS gyroscope calibration. *Mechatronics* 56 (2018), 242–253.

[6] Romit Roy Choudhury. 2021. Earable Computing: A New Area to Think About. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications*. 147–153.

[7] John L Crassidis, F Landis Markley, and Yang Cheng. 2007. Survey of nonlinear attitude estimation methods. *Journal of guidance, control, and dynamics* 30, 1 (2007), 12–28.

[8] Andrea Ferlini, Alessandro Montanari, Andreas Grammenos, Robert Harle, and Cecilia Mascolo. 2021. Enabling In-Ear Magnetic Sensing: Automatic and User Transparent Magnetometer Calibration. In *2021 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 1–8.

[9] Andrea Ferlini, Alessandro Montanari, Cecilia Mascolo, and Robert Harle. 2019. Head motion tracking through in-ear wearables. In *Proceedings of the 1st International Workshop on Earable Computing*. 8–13.

[10] Hassen Fourati. 2014. Heterogeneous data fusion algorithm for pedestrian navigation via foot-mounted inertial measurement unit and complementary filter. *IEEE Transactions on Instrumentation and Measurement* 64, 1 (2014), 221–229.

[11] Iuri Frosio, Federico Pedersini, and N Alberto Borghese. 2012. Autocalibration of triaxial MEMS accelerometers with automatic sensor model selection. *IEEE Sensors Journal* 12, 6 (2012), 2100–2108.

[12] Andreas Grammenos, Cecilia Mascolo, and Jon Crowcroft. 2018. You are sensing, but are you biased? a user unaided sensor calibration approach for mobile sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–26.

[13] Antonio R Jimenez, Fernando Seco, Carlos Prieto, and Jorge Guevara. 2009. A comparison of pedestrian dead-reckoning algorithms using a low-cost MEMS IMU. In *2009 IEEE International Symposium on Intelligent Signal Processing*. IEEE, 37–42.

[14] Justin Kavanagh, Rod Barrett, and Steven Morrison. 2006. The role of the neck and trunk in facilitating head stability during walking. *Experimental brain research* 172, 4 (2006), 454.

[15] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Allesandro Montanari. 2018. Earables for personal-scale behavior analytics. *IEEE Pervasive Computing* 17, 3 (2018), 83–89.

[16] Manon Kok, Jeroen D Hol, and Thomas B Schön. 2017. Using inertial sensors for position and orientation estimation. *arXiv preprint arXiv:1704.06053* (2017).

[17] Chuanhua Lu, Hideaki Uchiyama, Diego Thomas, Atsushi Shimada, and Rinichiro Taniguchi. 2019. Indoor positioning system based on chest-mounted IMU. *Sensors* 19, 2 (2019), 420.

[18] Dong Ma, Andrea Ferlini, and Cecilia Mascolo. 2021. OESense: Employing Occlusion Effect for In-ear Human Sensing. *arXiv preprint arXiv:2106.08607* (2021).

[19] Sebastian Madgwick. 2010. An efficient orientation filter for inertial and inertial/magnetic sensor arrays. *Report x-io and University of Bristol (UK)* 25 (2010), 113–118.

[20] Robert Mahony, Tarek Hamel, and J-M Pflimlin. 2005. Complementary filter design on the special orthogonal group SO (3). In *Proceedings of the 44th IEEE Conference on Decision and Control*. IEEE, 1477–1484.

[21] Anh Nguyen, Raghda Alqurashi, Zohreh Raghebi, Farnoush Banaei-Kashani, Ann C Halbower, and Tam Vu. 2016. A lightweight and inexpensive in-ear sensing system for automatic whole-night sleep stage monitoring. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*. 230–244.

[22] Stefanie Passler, Niklas Müller, and Veit Senner. 2019. In-ear pulse rate measurement: a valid alternative to heart rate derived from electrocardiography? *Sensors* 19, 17 (2019), 3641.

[23] James AC Patterson, Douglas C McIlwraith, and Guang-Zhong Yang. 2009. A flexible, low noise reflective PPG sensor platform for ear-worn heart rate monitoring. In *2009 sixth international workshop on wearable and implantable body sensor networks*. IEEE, 286–291.

[24] Jay Prakash, Zhijian Yang, Yu-Lin Wei, and Romit Roy Choudhury. 2019. Stear: Robust step counting from earables. In *Proceedings of the 1st International Workshop on Earable Computing*. 36–41.

[25] Valérie Renaudin, Muhammad Haris Afzal, and Gérard Lachapelle. 2010. Complete triaxis magnetometer calibration in the magnetic domain. *Journal of sensors* 2010 (2010).

[26] Brent A Renfro, Miquela Stein, Nicholas Boeker, and Audric Terry. 2018. An analysis of global positioning system (GPS) standard positioning service (SPS) performance for 2017. *See https://www. gps. gov/systems/gps/performance/2014-GPS-SPS-performance-analysis. pdf* (2018).

[27] Siddharth Rupavatharam and Marco Gruteser. 2019. Towards in-ear inertial jaw clenching detection. In *Proceedings of the 1st International Workshop on Earable Computing*. 54–55.

[28] Sheng Shen, Mahanth Gowda, and Romit Roy Choudhury. 2018. Closing the gaps in inertial motion tracking. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. 429–444.

[29] Martin Sipos, Pavel Paces, Jan Rohac, and Petr Novacek. 2011. Analyses of triaxial accelerometer calibration algorithms. *IEEE Sensors Journal* 12, 5 (2011), 1157–1165.

[30] Isaac Skog and Peter Händel. 2006. Calibration of a MEMS inertial measurement unit. In *XVII IMEKO world congress*. Citeseer, 1–6.

[31] Joan Sola. 2017. Quaternion kinematics for the error-state Kalman filter. *arXiv preprint arXiv:1711.02508* (2017).

[32] JL Weston and DH Titterton. 2000. Modern inertial navigation technology and its application. *Electronics & Communication Engineering Journal* 12, 2 (2000), 49–64.

[33] Seong-hoon Peter Won and Farid Golnaraghi. 2009. A triaxial accelerometer calibration method using a mathematical model. *IEEE transactions on instrumentation and measurement* 59, 8 (2009), 2144–2153.
[34] Oliver J Woodman. 2007. *An introduction to inertial navigation.* Technical Report. University of Cambridge, Computer Laboratory.
[35] Haotian Yang, Bin Zhou, Lixin Wang, Haifeng Xing, and Rong Zhang. 2018. A novel tri-axial MEMS gyroscope calibration method over a full temperature range. *Sensors* 18, 9 (2018), 3004.
[36] AS Zaidi and MR Suddle. 2006. Global navigation satellite systems: a survey. In *2006 international conference on advances in space technologies*. IEEE, 84–87.