EarMag: In-Ear Magnetosensing for Jaw and Head Gesture-Based Human-Computer Interaction

Max J. F. van Oort
Delft University of Technology
Delft, The Netherlands
m.j.f.vanoort@student.tudelft.nl

Gabriel E. Sáenz
Jawsaver BV
Utrecht, The Netherlands
gabriel@getsovn.com

Selina Tirtajana
Jawsaver BV
Utrecht, The Netherlands
selina@getsovn.com

Przemysław Pawełczak
Delft University of Technology
Delft, The Netherlands
p.pawelczak@tudelft.nl

Abstract

This study explores the use of in-ear magnetosensing (EarMag) as a novel sensing technique for detecting jaw and head movements in the context of human-computer interaction. While prior work has explored the use of acoustic, inertial, and visual sensing, the potential of EarMag remains to be explored. As a proof of concept, 17 orofacial physiotherapy-related exercises were collected from 21 participants using EarMag-enabled earables. A soft voting ensemble of support vector machine and random forest achieved 76% accuracy for five exercises on an unseen test set of ten users. While individual anatomical differences pose challenges for generalization, this work highlights the potential of EarMag for applications such as assistive technologies, silent speech interfaces, and biosignal tracking.

CCS Concepts

• **Human-centered computing** → **Gestural input**; • **Hardware** → **Sensor devices and platforms**.

Keywords

In-Ear Magnetosensing, Earables, Jaw Head Gesture Recognition, Human-Computer Interaction

ACM Reference Format:
Max J. F. van Oort, Gabriel E. Sáenz, Selina Tirtajana, and Przemysław Pawełczak. 2025. EarMag: In-Ear Magnetosensing for Jaw and Head Gesture-Based Human-Computer Interaction. In *Companion of the 2025 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp Companion '25), October 12–16, 2025, Espoo, Finland*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3714394.3757254

1 Introduction

In recent years, the field of Human-Computer Interaction (HCI) has experienced significant advances in computing power, computer vision, and machine learning techniques [15]. Hands- and eye-free HCI can be established by placing sensors on different parts of the body including the ear, while the anatomical properties and location of the ear create the most unique sensing opportunities. Earables, earbuds with a set of sensors, are widely used to fill in the sensing opportunities [19]. Two of the most common themes for earable sensing to enable HCI are head gestures and mouth gestures (including jaw, tongue, and teeth gestures) [16].

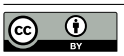
So far, head gestures have been mostly detected from inertial sensors within an earable. More specifically, accelerometer and gyroscope data [8, 12], but in-ear pressure sensing [2] has also been suggested. Jaw gestures can also be detected using in-ear pressure sensing [2]. In addition, proximity sensors [4] and piezoelectric bending sensors [6] have been researched for HCI.

For earable research, there are multiple ear-based sensing devices that offer (part of) the sensors mentioned above. The eSense [11] platform has been the most used platform for earable research. More recently, the OmniBuds [14] platform was released as a successor to eSense. The most recently released platform is the OpenEarable 2.0 [17], which offers, among others, a 9-axis Inertial Measurement Unit (IMU) located at the concha, a 3-axis IMU located at the ear canal entrance and a barometer located at the ear canal entrance.

What all of the platforms have in common is that they do not provide a magnetic sensor located at the ear canal. Currently, some platforms, such as OpenEarable 2.0, offer a 3-axis magnetic sensor located at the concha but not in the ear canal. The reason for this is that magnetic sensors have been found to be problematic due to the magnetic interference from the speakers [17]. However, these issues can be reduced with calibration methods [7] or by using piezoelectric speakers [9].

Still, magnetic sensors are only available in earables located at the concha and are mainly used for navigational purposes [7, 10]. To the best of our knowledge, current research has not explored the use of electromagnetic sensing in the ear canal for the detection of head and jaw movements for the purpose of HCI.

Recently developed earables (SOVN earbuds, Jawsaver BV, The Netherlands) use magnetic sensing of the external ear canal. The In-Ear Magnetosensing (EarMag) sensor is capable of measuring deformations of the ear canal ranging from the smallest deformations caused by heartbeat to larger deformations caused by jaw movements. More specifically, the condyle, as part of the Temporomandibular Joint (TMJ) [22], is deforming the ear canal when moving the jaw. This leads to the main question: *"How feasible is EarMag for detecting jaw and head movements to enable HCI?"*. By



UbiComp Companion '25, Espoo, Finland

ACM ISBN 979-8-4007-1477-1/2025/10
https://doi.org/10.1145/3714394.3757254

Table 1: Overview of in-ear sensing methods for jaw and head gesture recognition.

Study/ System	Sensor Type	Target Gestures	Method/ Model	Performance
CanalSense [2]	Barometer (air pressure)	Jaw (open/close), Head (nod, shake, tilt)	Random Forest	88% Jaw, 85% Head
IR Proximity [4]	3D IR Proximity Sensors	Jaw (open/close)	—	—
Piezo Sensor [6]	Piezoelectric bending strip	Jaw (open/close)	—	—
Laporte et al. [12]	3-axis Acc + Gyro (eSense)	Head (nodding, shaking)	Deep CNN	82% (F1)
Gashi et al. [8]	3-axis Acc + Gyro (eSense)	Head (nodding, shaking)	Hierarchical DNN + TL	88% (F1)
Groovemeter [13]	3-axis Gyroscope (eSense)	Head (music response)	LSTM (RNN)	81% (F1)

answering this question, the main contributions of this work are as follows:

- We present a novel dataset compiled as a part of this study, which contains the data of 17 orofacial physiotherapy related exercises collected from 21 participants.
- We show the potential of using EarMag to detect head and jaw movements to enable HCI by achieving 80% precision and 76% recall in five different movements of ten participants excluded from the training set.

A more comprehensive exploration of the methods and results presented in this paper is available in the corresponding master's thesis [20] (thesis is also available upon request to the first author). The source code and data are publicly available online [21].

2 Related Work

The published articles relate to in-ear sensing methods for the detection of jaw and head gestures. Table 1 provides a summary of the previously published methods.

2.1 Jaw Gestures

Jaw gestures can be detected from deformations in the ear canal caused by movement of the jaw. Specifically, movement of the condyle, part of the TMJ, alters the shape of the ear canal.

CanalSense [2] is a face-related movement recognition system based on air pressure sensing with an embedded barometer. Using commercially available earbuds customized with an embedded barometer and an Random Forest (RF) model, CanalSense can detect gestures involving sliding the jaw left and right, and opening and closing of the mouth with an accuracy of 88% across 12 participants. It can also detect four stages of opening the mouth (closed, slightly open, open, fully open) with an accuracy of 80%.

In contrast, proximity sensors are proposed to measure the deformation of the ear canal caused by movement of the lower jaw [4]. More precisely, three orthogonal infrared proximity sensors attached to a custom-fitted earpiece have been implemented to classify different jaw gestures, such as opening and closing of the mouth.

Lastly, a piezoelectric bending sensor is also suggested for the detection of jaw gestures [6]. The bending sensor consists of a thin piezoelectric strip attached to a custom-fitted earpiece and was used to detect jaw gestures such as opening and closing of the mouth.

2.2 Head Gestures

While CanalSense [2] was used for the detection of four jaw gestures, it was also designed to detect head gestures that involve facing left and right, facing up and down, and tilting the head left and right. CanalSense, using pressure sensing, was able to detect the head gestures with an accuracy of 85% across 12 participants.

Head gestures can also be detected using inertial sensors. Laporte et al. [12] used the eSense earables with a built-in 3-axis accelerometer and a 3-axis gyroscope for the detection of head shaking and nodding. The paper showed the potential of using an end-to-end deep convolutional neural network by achieving a F1-score of 82% based on ten participants.

Similarly, Gashi et al. [8] also used the eSense earables with the 3-axis accelerometer and 3-axis gyroscope for the detection of head shaking and nodding. In contrast, this work collected data from 21 participants, used a hierarchical approach based on deep neural networks, and used transfer learning that involves existing human activity recognition datasets. This research reported an F1-score of 88% on the head shaking and nodding exercises.

Lastly, Groovemeter [13] detects head motion reactions to music using the eSense earables with a 3-axis gyroscope. Long short-term memory, a type of recurrent neural network, was used to classify head motions collected from 30 participants. The evaluation showed an F1-score of 81% for head motion reactions with leave-one-subject-out cross-validation.

In contrast to these works, this work shows how EarMag-enabled earables can be used to recognize orofacial physiotherapy-related exercises, including jaw and head gestures, and breathing and massaging exercises.

3 A Need for Jaw and Head Gesture Recognition

There is a growing need for reliable jaw and head gesture recognition, particularly in the context of orofacial physiotherapy. Patients recovering from conditions such as TMJ disorders perform specific exercises that involve subtle jaw or head movements. Accurate gesture detection allows for progress tracking, remote supervision, and enhanced adherence while patients continue performing these exercises at home. While broader applications exist, such as assistive technologies, silent speech interfaces, and biosignal tracking, this work focuses on physiotherapy-related use cases.

Gesture recognition modalities based on IMUs, pressure sensors, and proximity sensors face limitations in detecting subtle gestures or localized movements and are mostly targeted to either jaw gestures or head gestures. In contrast, EarMag can be used to detect both jaw and head gestures due to the orientation and location of the magnet and the magnetosensor. Especially the in-ear positioning of the magnet provides more detail and makes EarMag able to detect more subtle deformations of the ear canal.

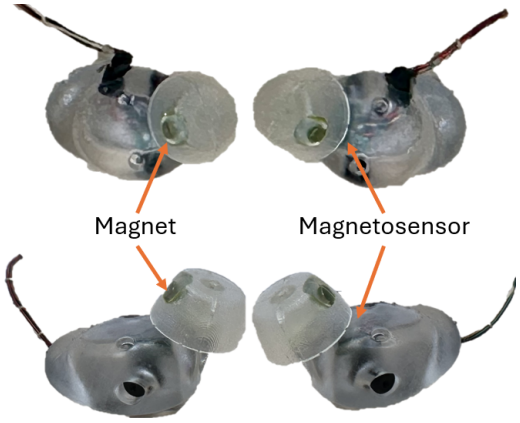


Figure 1: Top and side view of the EarMag-enabled earables used for the experiments. The earables are wired to a processing unit, which is not shown in the figure.

4 Study Design

4.1 EarMag Hardware Description

In this study, a pair of earables equipped with EarMag sensors is used for tracking jaw and head movements. The use of magnetic sensors in earables is already quite common in devices available on the market, primarily due to their small size, ease of integration, and low power consumption. The in-ear portion of the earables consists of a magnetic sensor housed in the body of the earpiece and a flexible silicon tip that is embedded with a small (~2 mm) rare earth disc magnet. Figure 1 shows the position of the magnet and the magnetosensor. As jaw movements deform the ear canal, the magnetized silicon tip moves in conjunction with this deformation. Such movements cause a relative displacement between the magnet in the silicon tip and the sensor, resulting in variations in magnetic flux density. These variations are measured by the magnetic sensor along three spatial axes (X, Y, Z), corresponding to jaw protrusion, opening/closing, and lateral movements, respectively.

Deformations of the magnet inside the silicon tip can also be caused by smaller movements, such as heartbeat and breathing. More specifically, heartbeat from the superficial temporal artery proximal to the ear canal causes micro movements of the magnetized tip, which can be captured by the magnetosensor.

This new method of biosensing allows for high-resolution, unobtrusive tracking of jaw activity, supporting applications such as bruxism detection, sleep monitoring, eating behaviour analysis, biosignal monitoring, and hands-free interaction using jaw gestures. The biosensor compatibility and ease of integration with standard in-ear wearables (e.g. True Wireless Stereo earbuds) further enhances its practicality and potential for widespread adoption. Further details regarding jaw movement tracking and biosensing using EarMag can be found in the patent application EP4440414A1 [18].

4.2 Participant Selection

To perform a structured data collection, a protocol was designed. The data collection protocol was approved by the Human Research Ethics Committee of the TU Delft. Before the participants were recruited, it was decided that the recruitment of volunteers would be random, so there were no restrictions in age, gender or other personal information. Eventually, 21 people, 12 males and nine females, participated in this research.

4.3 Data Collection

Before starting the data collection, each participant received and signed an informed consent form. Once consent was obtained, data collection could begin.

Cleaned earables were placed in the participants' ears, and each participant was seated in front of a monitor displaying a custom graphical user interface. The first screen showed a randomly generated Universal Unique Identifier (UUID) assigned to the participant. All captured data was stored anonymously under this UUID, known only to the lead researcher.

Next, a welcome screen briefly introduced the study's goal, followed by an instruction screen containing both textual and video explanations of the upcoming exercise. During this time, the lead researcher monitored the session and was available to answer any questions.

Once the participant confirmed understanding, they proceeded to the exercise screen. Most exercises consisted of three phases: performing the action within two seconds (e.g. opening the mouth), returning to a relaxed state in two seconds (e.g. closing the mouth), and pausing for one second before the next repetition. The action to be performed was highlighted in green and a short sound indicated transitions, which is particularly helpful for exercises involving head movements where the screen may not be visible. After eight iterations, the next exercise's instruction screen appeared automatically. This procedure was repeated for all 17 exercises described in the section below.

4.4 List of Considered Exercises

(1) **Jaw exercises**: opening and closing the mouth, opening and closing the mouth while pressing the tongue against the roof of the mouth, moving the lower jaw left, moving the lower jaw right, moving the lower jaw forward, and moving the lower jaw backward.
(2) **Head exercises**: looking down, looking up, creating a 'double chin' by moving the head forward and backward, moving the right ear to the right shoulder, moving the left ear to the left shoulder, looking right, and looking left.
(3) **Other exercises**: humming, massaging the jaw muscles in forward rotating circles, massaging the jaw muscles in backward rotating circles, and breathing in four cycles (breathe in, hold, breathe out, hold).

4.4.1 Selected Exercises. Out of the 17 available exercises, five exercises were selected for the classification model: *opening-closing mouth* (OC), *looking up* (LU), *looking right* (LR), *massaging jaw forward* (MJ), and *breathing* (BR). This subset was chosen because the classification complexity needed to be reduced and enough variety remains between the movement patterns. Furthermore, these exercises are relatively easy to perform considering that they will be used in a classification problem within the context of orofacial physiotherapy.

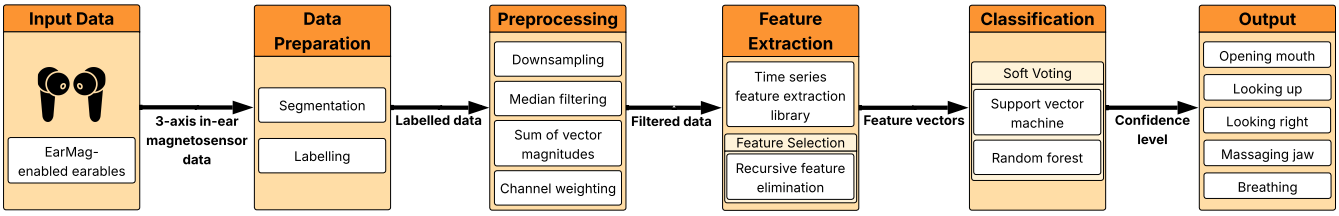


Figure 2: Processing pipeline of the EarMag system.

5 Methodology

Figure 2 shows the processing pipeline of the EarMag system. The methods used to implement each part of the system will be described below.

5.1 Input Data

While the participants were performing the exercises, EarMag data was collected with a sampling rate of approximately 196 Hz. This resulted in approximately 20 minutes of data per user from six different channels: three from the right earbud and three from the left earbud. The raw data, consisting of a timestamp and six channel values for each sample, were written to a *csv* file. Furthermore, a second file was generated for the raw annotations, consisting of the start and end timestamps, and the label of the exercise that was performed in that period of time.

5.2 Data Preparation

5.2.1 Segmentation. Segmentation of the data into shorter windows is needed, because the sensor data is recorded in the time domain. Knowing that the longest exercise duration could take up to four seconds (except for *massaging jaw* and *breathing*), window lengths larger than four seconds would only capture noise, so the following window lengths have been explored: [3.5, 3.75, 4.0] seconds. From testing, it appeared that the 3.75 second window would give the best results, so that window length was used. Another parameter for the segmentation was the overlap with the previous window, to ensure that no event was missed. From the set [50%, 65%, 75%, 80%] overlap, the 80% (three seconds) overlap gave the best results.

5.2.2 Labelling. After the segmentation was performed, each window should be assigned a label. However, this posed a slight challenge, because a window could contain multiple actions. This is because there is a one-second break in between the repetitions of the same exercise and a window could contain the end and the start of two sequential iterations. To solve this, the window should be labelled as the most dominant event inside that window. In this case, the most dominant event was declared to be at least 60% of the window samples.

5.3 Preprocessing

5.3.1 Downsampling. To remove high-frequency noise from the raw data, downsampling was applied. After analyzing the frequency spectra of the exercises, it was observed that the most dominant frequencies are below 4 Hz. Following Nyquist-Shannon's theorem, this translates to a minimum sample frequency of 8 Hz, but a small safety margin was built in to end up with a sample frequency of 10 Hz. Compared to the original 200 Hz data, this would result in a 95% decrease in data size, while it was observed that the performance did not decrease. Mean aggregation was applied as the downsampling method, where linear interpolation was used for missing values.

5.3.2 Median Filtering. Median filtering was implemented to remove sensor offsets and to remove the sensor drift that can be caused by Earth's magnetic fields or by slight changes in the ear tip position during exercises. A median filter with a window size of six seconds was used, which is larger than the duration of a single prediction window (3.75 seconds). This approach would result in a median that would be approximately equal to the signal baseline.

5.3.3 Sum of Vector Magnitudes. In time series analysis, multi-channel data is commonly combined into a single representative signal, to make the processing of the data easier and faster. However, when inter-channel dynamics or directionality matter, separate channel analysis is maintained. In this case it was assumed that the shape and magnitude of one combined signal should give sufficient information to differentiate the exercises. To combine the six channels into one signal, Sum Of Vector Magnitudes (SOVM) is used [5]: $\text{SOVM} = \sqrt{x_L^2 + y_L^2 + z_L^2} + \sqrt{x_R^2 + y_R^2 + z_R^2}$, where x_L, y_L and z_L represent the left earbud channels and x_R, y_R and z_R represent the right earbud channels. Figure 3 shows an example of the SOVM signal for each of the five selected exercises.

5.3.4 Channel Weighting. To improve the separability between similar exercises, channel weighting was explored as a final step in the preprocessing pipeline. The main idea was to increase the contrast between exercises by boosting channels that were strong for one gesture but weak for another.

This analysis focused on three exercises that were found to be the most similar: *opening-closing mouth*, *looking right*, and *looking up*. For each, the average amplitude across all repetitions for each channel was computed. Weighting factors were then derived to highlight distinct signal characteristics between exercises, in order to make the combined signal more discriminative.

Based on this analysis, channels x_L and x_R were assigned a higher weight of 1.25 in the final SOVM calculation for all exercises, while the remaining channels retained a weight of 1. This resulted in better separation for the most similar exercises while it did not have a negative impact on the other exercises.

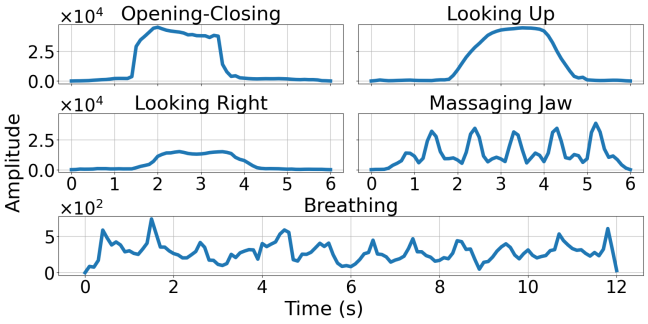


Figure 3: Example of filtered EarMag data of the five selected exercises for one user, where all figures show the combined SOVM signal (for definition see Section 5.3.3).

5.4 Feature Extraction and Selection

To prepare the filtered and weighted SOVM signals for classification, meaningful features need to be extracted. Afterwards, only a subset of features is selected based on feature importances.

Inspired by the feature extraction process of MedBuds [1], Time Series Feature Extraction Library (TSFEL) [3] is used. TSFEL is a Python package for efficient feature extraction from time series data. TSFEL automatically extracts 156 features that span the statistical, temporal, and spectral domains. Given the high amount of extracted features, feature selection was applied to remove redundant features and retain only the most important features. To select the most important feature, Recursive Feature Elimination (RFE) was used. It was chosen to use a linear Support Vector Machine (SVM) as a base estimator, because it is relatively fast and supports feature importances represented by coefficients. To speed up the RFE process, ten features were pruned after each round of feature elimination. This process was repeated until the desired number of features (K) remained, where $K \in \{5, 10, 15, 20, 25\}$ was investigated. The lowest number of features with the highest mean accuracy was chosen, which resulted in $K = 10$ selected features: (i) mean absolute differences, (ii) median absolute differences, (iii) signal distance, (iv) sum of absolute differences, (v) wavelet absolute mean at 2.5 Hz, (vi) wavelet energy at 0.42 Hz and 2.5 Hz, (vii) wavelet standard deviation at 0.28 Hz and 2.5 Hz, and (viii) spectral spread.

5.5 Classification

To evaluate the classification of the five selected exercises based on the extracted features, four widely used supervised machine learning algorithms were tested: SVM, logistic regression, RF, and extreme gradient boosting. In addition to evaluating these single classifiers, ensemble learning was also explored to potentially improve robustness. Specifically, a soft voting ensemble classifier was implemented in the evaluation pipeline, which combines the output probabilities of multiple individual classifiers. In soft voting, the output of every single classifier is a probability for every class, and the final prediction is based on the average of all output probabilities across all classifiers.

After testing ensembles up to three classifiers and comparing them with the single classifiers, it was decided to use an ensemble of SVM and RF as a soft classifier. It shared the best overall mean accuracy, but it also had the lowest standard deviation, which indicates that the classifier is the most stable. The combination of SVM and RF works particularly well due to their diversity. SVM is a margin-based, in this case linear, classifier that provides sharp boundaries. On the other hand, RF is a tree-based classifier that is more flexible and works well with non-linear data.

5.6 Evaluation Procedure and Validation

Five-fold cross-validation was chosen, because it provides a balance between computational efficiency and statistical reliability. In this case, with a dataset of 21 users, splitting it into five folds ensures that each fold will include four sets of four users and one set of five users. This setup will give the model unseen users during every fold, which will prepare the model for the real-world scenario. To evaluate the soft voting classifier, four different metrics will be used: precision, recall, specificity, and F1-score.

To assess the applicability and robustness of the trained ensemble model in real-time, a validation study was performed. The goal of the validation was to determine whether the trained model could correctly classify the selected exercises in a controlled setting.

A group of ten participants who were not present during training and testing of the model was recruited. This number of participants was chosen based on the size of the initial training and test dataset, which consisted of 21 users. Choosing around 50% of this set for validation was considered sufficient to evaluate generalizability.

5.7 Feature Weighting

To address the substantial variability in the performance of the exercises between users, individual feature weighting was applied. To find the optimal feature weights for each individual, Optuna, an open-source hyperparameter optimization framework, was used. Optuna performs efficient parameter searches by intelligently sampling hyperparameter combinations and pruning unpromising trials early. In this case, Optuna was configured to run 100 optimization trials, where the objective was to maximize both the precision and the accuracy of the predictions. To keep the search compact, the range of feature weights was restricted to $[0.75, 1.25]$, and choosing five windows to tune the feature weights resulted in the optimal balance between performance and the number of windows. In the future, individual feature tuning could be used as a calibration phase before real-time classification, where the user is asked to perform five repetitions of the selected exercises. As a result, each user would receive personalized weights for each of the ten selected features.

6 Results and Discussion

6.1 Individual Exercises

To understand the soft voting model's performance for each of the five selected exercises, an average confusion was made during training and testing of the model for the 21 participants. Figure 4 shows how *massaging jaw* (MJ) and *breathing* (BR) are relatively easy to predict compared to *looking right* (LR), where multiple false positives can be seen. Note that on average the rows add up to 84 instances of the exercises, so with 47 true positives out of the 84 instances, *looking right* is the clear bottleneck.

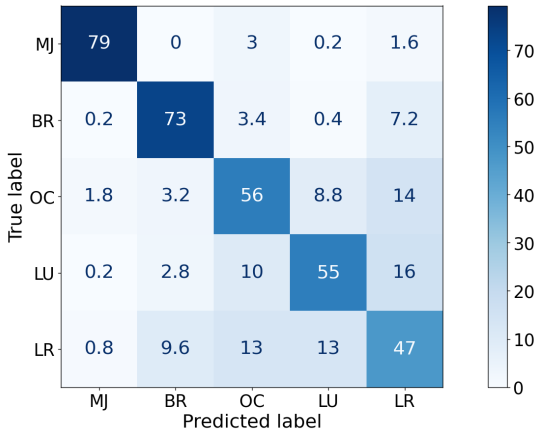


Figure 4: Average confusion matrix from five-fold cross-validation for 21 users for the five selected exercises (see Section 4.4.1).

Table 2: Classification results at various confidence thresholds for the five selected exercises for the ten user validation with feature weighting.

Threshold	Precision	Recall	Specificity	F1-score
0.2	78.6	77.5	94.4	75.6
0.3	78.6	76.9	94.4	75.3
0.4	80.3	75.5	95.1	75.2
0.5	80.6	68.3	97.0	71.3
0.6	79.4	56.7	98.8	62.4

6.2 Controlled Validation

We need to understand how the model would perform for unknown users. To find the optimal confidence threshold after the feature weighting process for the ten users, different confidence thresholds have been tested and shown in Table 2. With a precision of 80.3% and a recall of 75.5%, 0.4 is the optimal confidence threshold for real-time classification.

With an average precision and recall of around 60%, *looking right* is the hardest exercise to classify for the ensemble model. Based on these results, it should be noted that the average precision and recall of *looking right* are considered too low for a stable and user-friendly HCI system.

7 Limitations and Future Work

Traditional Learning. This project was limited to traditional machine learning methods, while deep learning algorithms might learn more complex features that would allow to differentiate similar jaw or head movements and improve the performance of the entire system.

Short Command Durations. The current command duration is limited to approximately two to four seconds, but a possible solution might be to change the segmentation and labelling process. Instead of labelling the entire exercise, labelling both the beginning and the end of the exercise separately would be more precise. This would remove the time constraint for the command duration, because the model would just wait for the ending label of the exercise after a start label was detected.

Delay Caused by Prediction Window. The limitation of waiting at least one prediction window to give a new command can be solved by a cool-down period. The cool-down period would start when a sequence of the beginning and end of the exercise is detected. With a cool-down period equal to the duration of the prediction window, the predictions on the transition windows in the cool-down period will be neglected.

Lack of Comparison to Other Methods. Another limitation is the lack of comparison to other in-ear methods of gesture detection. While our approach demonstrates promising results, it does not compare its efficacy and accuracy against existing methods and gestures used in previous HCI studies. Such comparisons are crucial as they provide a benchmark for evaluating the performance and usability of new technologies. Additionally, future research should compare power consumption and data & algorithm processing requirements for each method, as these factors are crucial considerations in embedded systems.

Sensor Fusion. Lastly, future research could include the exploration of sensor fusion. Only EarMag data provided by a 3-axis magnetic sensor has been considered in this project, but sensor fusion with a 3-axis accelerometer or gyroscope may improve the detection accuracy of the commands or increase the range of commands that can be detected by the earables.

8 Conclusions

The purpose of this work was to analyze whether the EarMag sensor data is feasible for the detection of jaw and head movements to enable HCI. The designed system was trained on EarMag data collected from 21 participants performing five orofacial physiotherapy related exercises: *opening-closing mouth*, *looking up*, *looking right*, *massaging jaw*, and *breathing*. With an average precision of 80%, recall of 76%, specificity of 95%, and F1-score of 75% at a confidence threshold of 0.4, the validation for ten unseen users shows promising results, but not high enough for HCI applications. After predicting the given commands in real-time during validation, it was found that the current setup works best when the gesture has a duration of 2-4 seconds with an interval of 3.75 seconds between gestures. Further research is needed to determine the most optimal system configuration to make EarMag feasible for gesture detection in HCI.

Acknowledgments

We would like to thank the rest of the SOVN team for their support, orofacial physiotherapist Simone Gouw for her help in selecting the 17 exercises and sharing clinical insights, and Vivian Dsouza for his early contributions.

References

[1] Murtadha Aldeer, David Waterworth, Zawar Hussain, Tahiya Chowdhury, Christian Brito, Quan Z. Sheng, Richard P. Martin, and Jorge Ortiz. 2022. MedBuds: In-Ear Inertial Medication Taking Detection Using Smart Wireless Earbuds. In *2022 2nd International Workshop on Cyber-Physical-Human System Design and Implementation* (Milan, Italy) *(CPHS '22)*. doi:10.1109/CPHS56133.2022.9804515

[2] 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)*. doi:10.1145/3126594.3126649

[3] Marília Barandas, Duarte Folgado, Letícia Fernandes, Sara Santos, Mariana Abreu, Patrícia Bota, Hui Liu, Tanja Schultz, and Hugo Gamboa. 2020. TSFEL: Time Series Feature Extraction Library. *SoftwareX* 11 (2020). doi:10.1016/j.softx.2020.100456

[4] 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. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers* (Osaka, Japan) *(UbiComp/ISWC'15 Adjunct)*. doi:10.1145/2800835.2807933

[5] Erika Bondareva, Elín Rós Hauksdóttir, and Cecilia Mascolo. 2021. Earables for Detection of Bruxism: a Feasibility Study. 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* (Virtual, USA) *(UbiComp/ISWC '21 Adjunct)*. doi:10.1145/3460418.3479327

[6] Johan Carioli, Aidin Delnavaz, Ricardo J. Zednik, and Jérémie Voix. 2018. Piezoelectric Earcanal Bending Sensor. *IEEE Sensors Journal* 18 (03 2018). doi:10.1109/JSEN.2017.2783299

[7] 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* (Kassel, Germany) *(PerCom '21)*. doi:10.1109/PERCOM50583.2021.9439112

[8] Shkurta Gashi, Aaqib Saeed, Alessandra Vicini, Elena Di Lascio, and Silvia Santini. 2021. Hierarchical Classification and Transfer Learning to Recognize Head Gestures and Facial Expressions Using Earbuds. In *Proceedings of the 2021 International Conference on Multimodal Interaction* (Montréal, QC, Canada) *(ICMI '21)*. doi:10.1145/3462244.3479921

[9] Chiara Gazzola, Valentina Zega, Fabrizio Cerini, Silvia Adorno, and Alberto Corigliano. 2023. On the Design and Modeling of a Full-Range Piezoelectric MEMS Loudspeaker for In-Ear Applications. *Journal of Microelectromechanical Systems* 32 (12 2023). doi:10.1109/JMEMS.2023.3312254

[10] Jian Gong, Xinyu Zhang, Yuanjun Huang, Ju Ren, and Yaoxue Zhang. 2021. Robust Inertial Motion Tracking through Deep Sensor Fusion across Smart Earbuds and Smartphone. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5 (06 2021). doi:10.1145/3463517

[11] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Alessandro Montanari. 2018. Earables for Personal-Scale Behavior Analytics. *IEEE Pervasive Computing* 17 (07 2018). doi:10.1109/MPRV.2018.03367740

[12] 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* (Virtual, USA) *(UbiComp/ISWC '21 Adjunct)*. doi:10.1145/3460418.3479322

[13] Euihyeok Lee, Chulhong Min, Jaeseung Lee, Jin Yu, and Seungwoo Kang. 2023. GrooveMeter: Enabling Music Engagement-aware Apps by Detecting Reactions to Daily Music Listening via Earable Sensing. In *Proceedings of the 31st ACM International Conference on Multimedia* (Ottawa ON, Canada) *(MM '23)*. doi:10.1145/3581783.3611968

[14] Alessandro Montanari, Ashok Thangarajan, Khaldoon Al-Naimi, Andrea Ferlini, Yang Liu, Ananta Narayanan Balaji, and Fahim Kawsar. 2024. OmniBuds: A Sensory Earable Platform for Advanced Bio-Sensing and On-Device Machine Learning. arXiv:2410.04775 [cs.ET] doi:10.48550/arXiv.2410.04775

[15] Kaushik Ranade, Tanmay Khule, and Riddhi More. 2024. Object Recognition in Human Computer Interaction:- A Comparative Analysis. arXiv:2411.04263 [cs.HC] doi:10.48550/arXiv.2411.04263

[16] Tobias Röddiger, Christopher Clarke, Paula Breitling, Tim Schneegans, Haibin Zhao, Hans Gellersen, and Michael Beigl. 2022. Sensing with Earables: A Systematic Literature Review and Taxonomy of Phenomena. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6 (09 2022). doi:10.1145/3550314

[17] Tobias Röddiger, Michael Küttner, Philipp Lepold, Tobias King, Dennis Moschina, Oliver Bagge, Joseph A. Paradiso, Christopher Clarke, and Michael Beigl. 2025. OpenEarable 2.0: Open-Source Earphone Platform for Physiological Ear Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9 (03 2025). doi:10.1145/3712069

[18] Gabriel Enrique Sáenz, Selina Tirtajana, David Benjamin Jaroch, and Joost Plattel. 2022. Jaw Movement Tracking System and Method. Patent Application EP4440414A1, 29 November 2022. https://patents.google.com/patent/EP4440414A1/

[19] Thomas Sepanosian and Ozlem Durmaz Incel. 2024. Boxing Gesture Recognition in Real-Time using Earable IMUs. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Melbourne VIC, Australia) *(UbiComp '24)*. doi:10.1145/3675094.3680524

[20] Max J. F. van Oort. 2025. *In-Ear Human-Computer Interaction: Using EarMag to Detect Head and Jaw Movements.* MSc thesis. Delft University of Technology, Delft, The Netherlands. https://resolver.tudelft.nl/uuid:824248db-24a0-4b9b-9628-d2aa81a1cb90

[21] Max J. F. van Oort, Gabriel E. Sáenz, Selina Tirtajana, and Przemysław Pawełczak. 2025. EarMag: In-Ear Magnetosensing for Jaw and Head Gesture-Based Human-Computer Interaction. https://github.com/jawsaver/earmag-ubicomp2025.

[22] Grzegorz Zieliński, Beata Pająk-Zielińska, and Michał Ginszt. 2024. A Meta-Analysis of the Global Prevalence of Temporomandibular Disorders. *Journal of Clinical Medicine* 13 (02 2024). doi:10.3390/jcm13051365