2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Hearables: Making Sense from Motion Artefacts in Ear-EEG for Real-Life Human Activity Classification

Ghena M. Hammour and Danilo P. Mandic

Abstract— **Ear-worn devices are rapidly gaining popularity as they provide the means for measuring vital signals in an unobtrusive, 24/7 wearable and discrete fashion. Naturally, these devices are prone to motion artefacts when used in out-of-lab environments, various movements and activities cause relative movement between user's skin and the electrodes. Historically, these artefacts are seen as nuisance resulting in discarding the segments of signal wherever such artefacts are present. In this work, we make use of such artefacts to classify different daily activities that include sitting, speaking aloud, chewing and walking. To this end, multiple classification techniques are employed to identify these activities using 8 features calculated from the electrode and microphone signal embedded in a generic multimodal in-ear sensor. The results show an overall training accuracy of 93% and 90% and a testing accuracy of 85% and 80% when using a KNN and a 2-layer neural network respectively, thus providing a much needed, simple and reliable framework for real-life human activity classification.**

I. INTRODUCTION

Wearable sensors are revolutionizing the way we acquire, analyse and interpret physiological data. These sensors have penetrated all aspects of life, from smart watches that monitor heart rate, oxygen saturation, activity level and water intake, to smart home assistants which can detect emotions from voice. Recent advances in this area have investigated ear-worn devices which are referred to as Hearables [1] and [2]. Benefiting from the privileged position of the head on a human body, Hearables have proven their capability in monitoring both physiological (electroencephalograph, electrocardiogram, heart rate, temperature, respiration), [3], [4], [5], [6], [7] and non-physiological (speech, user authentication) signals [8]. The ultimate aim of Hearables in both e-Health and recreational applications is to acquire useful data from the user in a 24/7 and unobtrusive fashion. The envisaged continuous operation will make it possible to provide insights into for example, health, activity during the day, quality of sleep at night and the overall well-being. However, continuous real-life monitoring out of the lab comes with the burden of the inevitable motion artefacts such as those due to speaking, chewing and walking. These artefacts heavily corrupt the signal of interest and more often than not, the epochs containing these artefacts are discarded leading to a loss and discontinuities in data.

To this end, we set out to make sense from the Ear-EEG artefacts and propose a novel method for human activity classification from motion artefacts induced through different routine daily activities. This is highly beneficial, not only from the viewpoint of "no data is bad data", but also given that human activity recognition has gained plenty of interest in e.g care for the elderly. Indeed, several systems have been designed with the aim of alleviating physical, emotional and economic burden of a caregiver or family member caring for an elderly with dementia [9], [10] with 152 million people expected to be affected by dementia by 2050 [11]. Lentzas and Vrakas in [12] reviewed the sensors, systems and methods used for activity recognition for the elderly. These include smart phones, wearable sensors (accelerometer, gyroscope, GPS, photoplethysmogram, temperature) and ambient sensors (pressure, humidity, infrared, magnetic switches, electric power usage). Such sensors are used either as a stand-alone or in combination and mostly achieve accurate results. However, with a view of adoption by a broader community including the elderly, most systems in the literature have one or more of the following drawbacks:

- Requirement for multiple wearable and non-wearable sensors and are built around an idea to be integrated into smart homes which makes them expensive;
- Assumption of a certain level of technological literacy for their operation – often prohibitive with the elderly;
- Most do not identify the user that performs an action.

On the other hand, Hearables represent a stand-alone device that is equipped with a multitude of physiological sensing functionality; such a solution can even identify users using their biometric data [8].

This pilot study takes the Hearables concept further and establishes a framework for making sense from artefacts by using them as a platform to infer behavioural queues. Human activity classification typically employs machine learning (ML) tools in standard settings; in [13], [14], [15], [16], the authors used ML tools for human activity classification and attained accuracies as high as 95%. In this study we conclusively demonstrate the possibility to classify the most frequent daily activities (sitting, speaking, chewing and walking) with a mean accuracy of 90%.

II. METHODS

A. Experimental setup and protocol

The Ear-EEG sensor used in this study represents a modified version of our original multi-modal generic EEG sensor used in [2], [17], whereby an electret condenser microphone (ECM) (model: CMC-4015-40L100) was embedded onto a

*This work was supported by the Dementia Research Institute (DRI).

G. Hammour and D.P. Mandic are with the Department of Electrical and Electronic Engineering, Imperial College London, SW7 2BT United Kingdom.
ghena.hammour17@imperial.ac.uk
d.mandic@imperial.ac.uk

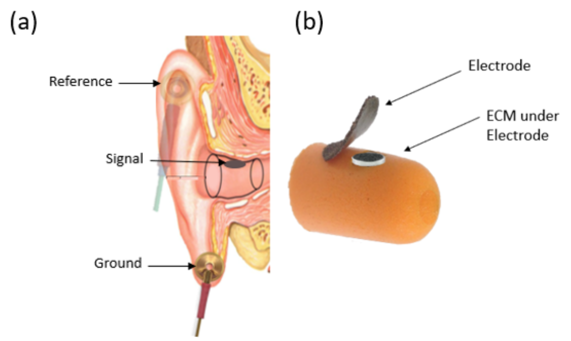

Fig. 1. Concept of a generic Ear-EEG multimodal sensor with a visible electrode and microphone. (a) Ear-EEG electrodes setup. (b) In-ear electro-mechanical sensor in the form of an electret microphone beneath a flexible electrode.

generic visco-elastic earplug, as shown in Fig. 1(b).
The Ear-EEG sensor has an electrode placed on the earpiece, and a conductive gel was applied before insertion into the ear. A reference gold-cup electrode was secured on the pinna of the ear, as shown in Fig. 1(a), and a ground gold-cup electrode was secured on the earlobe. The microphone signal and the Ear-EEG signals were simultaneously recorded using the g.tech g.USBamp amplifier with a 24-bit resolution and at a sampling rate of 1200 S/s – in accordance with previous settings in our studies [18].
The aim of our experiment was to investigate, study and classify the artefacts induced by different daily activities, namely, sitting, speaking, chewing and walking. For this pilot study, a 29-year old female subject was recorded over 6 different trials on different days. This setup was considered statistically sound according to [19]. During each trial, the subject was asked to insert a gelled in-ear sensor with the microphone embedded in it. The earlobe and helix were then abraded by an abrasive electrode gel and subsequently standard gold-cup electrodes were fixed to the respective positions by the means of medical tape. During the protocol, the subject sat in a home-setup on a comfortable chair facing a computer screen and was video recorded for time-stamping purposes. The subject was first asked to sit in a natural posture for 2 minutes; following that a text was displayed on the computer screen and a sound of a click was played to signal the subject to start reading out loud the displayed text for 2 minutes. This was then followed by a 30 second break where the subject sat without motion in a natural posture. Next, the subject was handed a chewing gum and chewed for 2 minutes, after which a 30 second break was re-initiated. Finally, the subject stood and walked on the spot for 2 minutes and rested for 30 seconds. The daily activities protocol (sitting, speaking, chewing, walking) was repeated twice for each trial. The recordings were performed under the IC ethics committee approval JRCO 20IC6414, and the subject gave full informed consent.

B. Data pre-processing and preparation

The recorded data from the 6 different trials were first time-stamped based on the protocol and the video recording acquired. Each activity was then appropriately labelled (sitting/ speaking/ chewing/ walking). The data were then pre-processed through the following steps:

- Mean removal and filtering through a 0.5-45Hz 8th order Butterworth filter;
- Normalization by dividing by the maximum value of the signal;
- Extraction of different activities from each trial based on the protocol and video time-stamping;
- Concatenation of same activities from different trials;
- Segmentation of the signals into 3-second segments and removal of segments that contain electrical noise;
- Division of data into 70% training data to determine the best set of features and train different classification models, and 30% testing data unseen by the model.

This resulted in a balanced data set summarized in Table I (772 observations in total, each observation was 3 seconds long) divided into 70% for training and 30% testing that was used. Figure 2 shows a labelled sample segment of pre-processed data from the electrical and mechanical recording modalities.

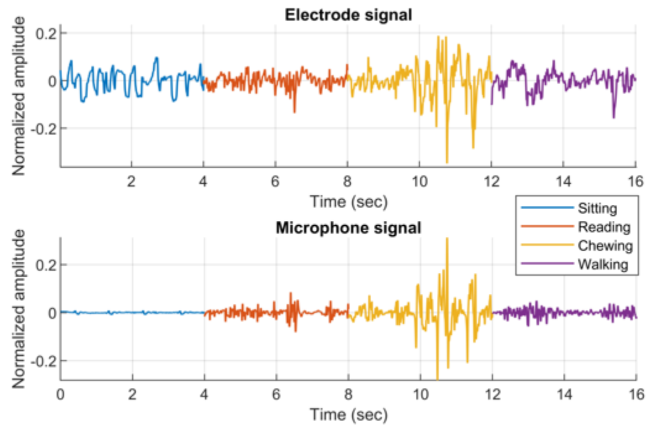

Fig. 2. Sample segments of pre-processed data from the electrode (top) and the microphone signals (bottom).

TABLE I
ACTIVITIES, CLASSES AND NUMBER OF OBSERVATIONS PER CLASS.

Classes	0: Sitting 1: Speaking 2: Chewing 3: Walking
No. of training observations	135 observations per class
No. of testing observations	58 observations per class

C. Feature selection and extraction

To extract the most important features corresponding to each class, the Control System Analysis and Design MATLAB Toolbox [20] was employed to calculate the commonly used time and frequency domain features, shown in Table II. The concatenated activity-specific observations of all trials were divided into 3-second epochs, and the time and frequency-domain features were calculated for both the electrode and microphone signal.

TABLE II

THE FEATURES CONSIDERED [20].

Type	Metric
Basic statistics	Mean Standard deviation RMS Shape factor
Higher order statistics	Kurtosis Skewness
Impulsive metrics	Peak value Impulse factor Crest factor Clearance factor
Signal processing metrics	Signal-to-Noise ratio (SNR) Total Harmonic Distortion (THD) Signal to Noise & Distortion ratio (SINAD)
Spectral metrics	Peak amplitude Peak frequency Band Power

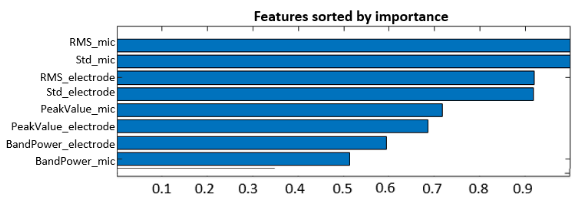

Fig. 3. Calculated features sorted by importance based on one-way ANOVA test.

In order to select the most significant features that allow for maximum separation between different classes and most affect the target output, the one-way analysis of variance (ANOVA) test was used. The choice of this method was suggested by the type of data, whereby the input variables (features) are numerical and the output/target variables (classes) are categorical [21]. The one-way ANOVA results were then rank ordered based on the significance of features. Only the features with significance of more than 50% from each recording modality were selected. It is important to note that the feature selection process was performed on the training data only. The selected features, shown in Fig. 3, were then used to train a comprehensive set of classifiers, and also for the test data.

D. Classifiers training & selection

Different classification model types, listed in Table III, were evaluated using the Machine Learning and Deep Learning MATLAB Toolbox [22]. Five-fold cross-validation was used on the randomized training data, and the classifier with the highest model accuracy was chosen. To evaluate the performance on unseen data, the overall model and class-specific accuracies were used as metrics to evaluate the performance of the classifier. In addition, the sensitivity (true positive rate (TPR)), false negative rate (FNR), precision (positive predictive rate (PPV)), and the false discovery rate

TABLE III

THE DIFFERENT TYPES OF CLASSIFIERS TRAINED

Model Type	Classifier
Decision Trees	Fine Tree (Max no. of splits: 100) Medium Tree (Max no. of splits: 20) Coarse Tree (Max no. of splits: 4)
Discriminant Analysis	Linear Discriminant Quadratic Discriminant
Naive Bayes Classifiers	Gaussian Naive Bayes Kernel Naive Bayes
Support Vector Machines (SVM)	Linear SVM Quadratic SVM Cubic SVM Fine SVM Medium SVM Coarse SVM
Nearest Neighbors Classifiers	Fine KNN Medium KNN Coarse KNN Cosine KNN Cubic KNN Weighted KNN
Ensemble Classifiers	Boosted trees Bagged trees Subspace Discriminant Subspace KNN RUSBoosted Trees

(FDV) were calculated as:

$$TPR = \frac{TP}{TP + FN} \times 100$$
$$FNR = \frac{FN}{FN + TP} \times 100$$
$$PPV = \frac{TP}{TP + FP} \times 100$$
$$FDR = \frac{FP}{FP + TP} \times 100$$

E. Neural network classifiers

A 2-layer feed-forward neural network, shown in Fig. 4, was used to classify the features from Table II into one of the targets (sitting, speaking, chewing, walking). One of the main challenges was deciding on the number of hidden layers and the number of neurons in each layer. As there exists no analytical solution or rule for their determination, these hyper-parameters were chosen based on systematic experimentation depending on the application at hand. Another challenge that naturally arises when training a neural network is the varying accuracy when training on the same data. This is mainly due to the random shuffling of data and the random initializations of the weights in the network. While this gives neural networks the flexibility to better learn, it also poses a problem of instability between subsequent training trials.

To address these issues, exhaustive search on the number of neurons in the two layers was performed to determine the optimal number of neurons in a neural network architecture. Each combination was repeated over 100 trials (using same seeds in order to compare different models) to address the variable accuracy results in subsequent training sessions. The mean overall accuracy of each combination was calculated

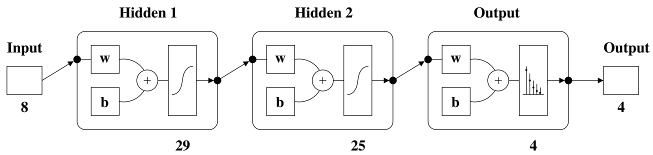

Fig. 4. Architecture of the neural network used

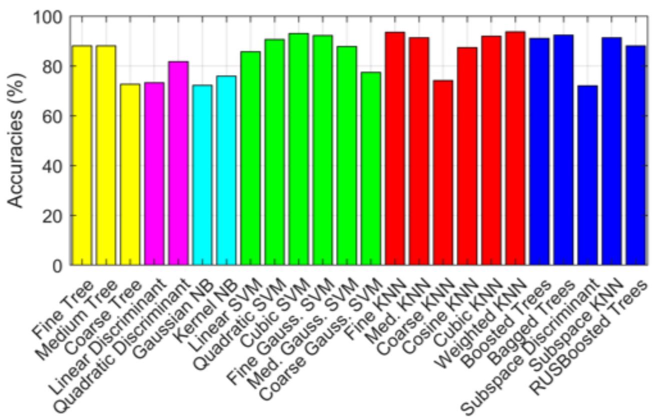

Fig. 5. Model accuracy of the different trained classifier models.

and the data was tested and reported for the models which yielded accuracy higher than 75%.

III. Results and Discussion

Given the proof-of-concept nature of this study, and the fact that EEG vastly varies between different trials, even from the same subject, our hypothesis was that the artefacts will be less varying and therefore usable for identification and classification. To this end, we evaluated the performance of a comprehensive set of standard and advanced classifiers listed in Table III. Fig. 5 illustrates the performance of the considered classifiers, with the KNN and neural network classifiers achieving best results and evaluated below.

A. K-nearest neighbour (KNN) classifier

The KNN classifier with k=10 had the highest overall accuracy. The distance metric used was the Euclidean distance, with the squared inverse as a distance weighting function. The overall trained model accuracy was high around 93.7%. This model was evaluated on the testing data and the overall testing accuracy was around 85%, while the overall sensitivity was 85% and the overall precision 87.75%. This means that on average, when the result is positive, our model does detect it as positive but also in a precise way, that is, when it predicts a correct class, that class is actually correct around 88% of the time. However, sensitivity drastically drops to 62.1% in the case of walking activity, as shown in Fig. 6, which in turn affects the precision of predicting the chewing activity, given that on several instances it was misclassified as walking.

B. Neural network classifier

Fig. 7 presents the trained model accuracy as a function of the number of neurons in the two layers. Note that, as

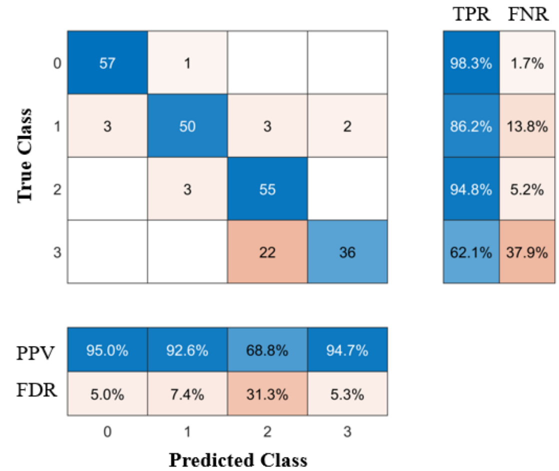

Fig. 6. Confusion matrix for the unseen data in the four-category prediction task with the activities 0: Sitting, 1: Speaking, 2: Chewing, and 3: Walking.

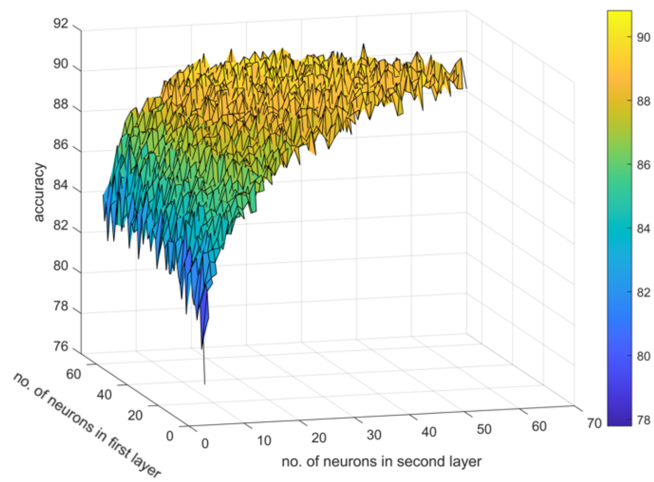

Fig. 7. Model accuracy of a trained neural network as a function of the number of neurons in each layer.

commonly suggested in the literature, when determining the number of neurons in subsequent hidden layers, their number in the second layer was kept equal to or smaller than that of the first layer. From the steep observed slope in Fig. 7, we deduce that a network with a small number of neurons was more sensitive to their number in the second layer. In fact, an increase from 5 to 10 neurons in the second layer enhanced model accuracy from around 81.5% to 85.5% (with the number of neurons in the first layer kept at 10), an improvement of 4%. On the other hand, doubling the number of neurons from 5 to 10 in the first layer resulted in a slight improvement from 81% to 81.5% (with the number of neurons in the second layer set to 5).

Table IV shows the mean test accuracy for different combinations of neuron population in the layers. Observe that performance improvement was only marginal for 25+ neurons in each layer. Therefore, and given the computational benefits, a 2-layer feed-forward neural network with 29

TABLE IV
NEURAL NETWORK MODEL ACCURACY FOR DIFFERENT COMBINATIONS OF THE NUMBERS OF NEURONS IN NN LAYERS.

No. of neurons in 1st layer	No. of neurons in 2nd layer	Mean model accuracy %	Mean test accuracy %
4	4	77.79	61.40
12	9	85.79	71.21
18	17	89.96	74.19
29	25	89.05	78.14
40	36	89.91	79.64
61	**43**	**90.82**	**80.40**
70	35	90.55	80.21

neurons in the first layer and 25 in the second layer was deemed adequate.

IV. CONCLUSIONS

We have proposed a radically new approach for human activity classification based on artefacts from the Ear-EEG modality from Hearables. The recording and analysis have conclusively indicated that artefact-corrupt epochs of EEG, which are usually discarded, represent a rich behavioural information stream that provides a new tool for real-life continuous behavioural monitoring. The analysis over a comprehensive sample of classifiers, has shown that a KNN classifier attained an overall classification accuracy of 85%, with a an overall precision of 87.5%. For rigour, the model was trained on data acquired on different days and the performance was tested on unseen data. This has also conclusively demonstrated our hypothesis that the artefacts exhibit a more regular behaviour than the EEG, when recorded over several days and multiple trials. It is our hope that this proof-of concept will lay the foundation for larger studies whereby EEG paradigms may be designed to be more ”user-specific” and the recording artefacts more generalizable. Our ongoing work examines the concept of transfer learning in this context, thus eliminating the need to retrain the model.

REFERENCES

[1] D. Looney, P. Kidmose, C. Park, M. Ungstrup, M. Rank, K. Rosenkranz, and D. Mandic, “The in-the-ear recording concept: User-centered and wearable brain monitoring,” *IEEE Pulse*, vol. 3, no. 6, p. 32–42, 2012.

[2] V. Goverdovsky, W. von Rosenberg, T. Nakamura, D. Looney, D. J. Sharp, C. Papavassiliou, M. J. Morrell, and D. P. Mandic, “Hearables: Multimodal physiological in-ear sensing,” *Scientific Reports*, vol. 7, no. 1, pp. 6948–10, 2017.

[3] G. Hammour, M. Yarici, W. v. Rosenberg, and D. P. Mandic, “Hearables: Feasibility and validation of in-ear electrocardiogram,” in *Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5777–5780, 2019.

[4] H. J. Davies, I. Williams, N. S. Peters, and D. P. Mandic, “In-ear SpO2: A tool for wearable, unobtrusive monitoring of core blood oxygen saturation,” *Sensors*, vol. 20, no. 17, pp. 1424–8220, 2020.

[5] T. Nakamura, T. Adjei, Y. Alqurashi, D. Looney, M. J. Morrell, and D. P. Mandic, “Complexity science for sleep stage classification from EEG,” in *Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 4387–4394, 2017.

[6] V. Goverdovsky, D. Looney, P. Kidmose, and D. P. Mandic, “In-ear EEG from viscoelastic generic earpieces: Robust and unobtrusive 24/7 monitoring,” *IEEE Sensors Journal*, vol. 16, no. 1, pp. 271–277, 2016.

[7] K. Matsumoto, Y. Temiz, H. Taghavi, E. L. Cornelius, H. Mori, and B. Michel, “An earbud-type wearable (a hearable) with vital parameter sensors for early detection and prevention of heat-stroke,” *In Proceedings of the Annual IEEE Conference Eng Med. Biol. Soc.*, pp. 7049–7055, Jul 2019.

[8] T. Nakamura, V. Goverdovsky, and D. P. Mandic, “In-ear EEG biometrics for feasible and readily collectable real-world person authentication,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 648–661, 2018.

[9] J. Blott, “Smart homes for the future of dementia care,” *Lancet neurology*, vol. 20, no. 4, pp. 264–264, 2021.

[10] S. Enshaeifar, A. Zoha, A. Markides, S. Skillman, S. T. Acton, T. Elsaleh, M. Hassanpour, A. Ahrabian, M. Kenny, S. Klein, H. Rostill, R. Nilforooshan, and P. Barnaghi, “Health management and pattern analysis of daily living activities of people with dementia using in-home sensors and machine learning techniques,” *PloS one*, vol. 13, no. 5, pp. 0195605–0195605, 2018.

[11] C. Su, L. Fu, Y. Chien, and T. Li, “Activity recognition system for dementia in smart homes based on wearable sensor data,” in *Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 463–469, 2018.

[12] A. Lentzas and D. Vrakas, “Non-intrusive human activity recognition and abnormal behavior detection on elderly people: A review,” *Artificial Intelligence Reviews*, vol. 53, no. 3, pp. 1975–2021, 2019.

[13] A. B. Mesanza, S. Lucas, A. Zubizarreta, I. Cabanes, E. Portillo, and A. Rodriguez-Larrad, “A machine learning approach to perform physical activity classification using a sensorized crutch tip,” *IEEE Access*, vol. 8, pp. 210023–210034, 2020.

[14] M. N. Ahmadi, T. G. Pavey, and S. G. Trost, “Machine learning models for classifying physical activity in free-living preschool children,” *Sensors*, vol. 20, no. 4364 1–14, 2020.

[15] A. Mannini and A. M. Sabatini, “Machine learning methods for classifying human physical activity from on-body accelerometers,” *Sensors*, vol. 10, no. 2, p. 1154–1175, 2010.

[16] S. Gandhi, *Artifact Detection and Removal from Electroencephalography (EEG) Data*. PhD thesis, University of Maryland, Baltimore County (UMBC), 2020.

[17] D. Looney, D. Mandic, and V. Goverdovsky, “Biosensing electrodes,” Nov. 2019. US patent US10463309B2.

[18] T. Nakamura, Y. D. Alqurashi, M. J. Morrell, and D. P. Mandic, “Hearables: Automatic overnight sleep monitoring with standardized in-ear eeg sensor,” *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 1, pp. 203–212, 2020.

[19] J. L. Ristić-Djurović, S. Ćirković, P. Mladenović, N. Romčević, and A. M. Trbovich, “Analysis of methods commonly used in biomedicine for treatment versus control comparison of very small samples,” *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 153–162, 2018.

[20] “Matlab control systems and analysis toolbox,” MATLAB R2020b. The MathWorks, Natick, MA, USA.

[21] K. J. Max Kuhn, *Applied Predictive Modeling*. Springer-Verlag New York, 8 ed., 2013.

[22] “Matlab machine learning and deep learning toolbox,” MATLAB R2020b. The MathWorks, Natick, MA, USA.