Gesture Recognition for Communication Support in the Context of the Bedroom: Comparison of Two Wearable Solutions

Ana Patrícia Rocha[1][a], Diogo Santos[2], Florentino Sánchez[2], Gonçalo Aguiar[2], Henrique Ramos[2], Miguel Ferreira[2], Tiago Bastos[2], Ilídio C. Oliveira[1][b] and António Teixeira[1][c]

[1] *Institute of Electronics and Informatics Engineering of Aveiro (IEETA), Department of Electronics, Telecommunications and Informatics, Intelligent Systems Associate Laboratory (LASI), University of Aveiro, Aveiro, Portugal*

[2] *Department of Electronics, Telecommunications and Informatics, University of Aveiro, Aveiro, Portugal*

{*aprocha, diogoejsantos, florentino, gonc.soares12, henriqueframos, miguelmcferreira, tiagovilar07, ico, ajst*}*@ua.pt*

Keywords: Gestures, Human Communication, Smart Environments, Bed, Wearable Sensors, Smartwatch.

Abstract: Gestures can be a suitable way of supporting communication for people with communication difficulties, especially in the bedroom scenario. In the scope of the AAL APH-ALARM project, we previously proposed a gesture-based communication solution for the bedroom context, which relies on a smartwatch for gesture recognition. In this contribution, our main aim is to explore better wearable alternatives to the smartwatch regarding the form factor and comfort of use, as well as cost. We compare a smartwatch and a simpler, smaller, less expensive wearable device from MbientLab, both integrating an accelerometer and a gyroscope, in terms of gesture classification performance. The results obtained based on data acquired from six subjects and the support vector machines algorithm show that, overall, both explored devices lead to a model with promising and similar results (mean accuracy and F1 score of 98%, and mean false positive rate of 2%), being thus possible to rely on a smaller and lower cost wearable device, such as the MbientLab sensor module, for recognizing the considered arm gestures.

1 INTRODUCTION

Verbal communication plays an important role in our lives, allowing us to express ourselves to others (Love and Brumm, 2012). Therefore, when the ability to use language is affected, such as when a person acquires a language disorder (e.g., resulting from brain damage due to a stroke or a neurological disease) (Love and Brumm, 2012), it has a considerable negative impact on the person's life, leading for example to a loss of independence and sense of safety.

Communication is especially important in some daily life scenarios, such as the in-bed scenario (i.e., person lying in bed). In this scenario, which motivates the present research, a person with communication difficulties may be alone and need to ask for immediate help if they suddenly fell unwell. Furthermore, people that acquired a language disorder (e.g., due to a stroke or Parkinson's disease), may need help with more common situations, such as getting up from bed to go to the bathroom during the night, due to the fear of falling.

Although a large offer of augmentative and alternative communication (AAC) solutions is currently available for aiding people with communication difficulties, many of them rely on applications for mobile devices with touchscreens, in which the user selects words or sentences represented by pictograms and/or text to generate sentences which are transmitted to others using synthesized speech (Elsahar et al., 2019). Moreover, research on assistive communication has also focused mainly on mobile applications (Allen et al., 2007; Kane et al., 2012; Laxmidas et al., 2021; Obiorah et al., 2021).

While this type of solution may be adequate in many daily life situations, they may prove less practical for the in-bed scenario, since they require the user to move around to reach for the device or having to hold it in an uncomfortable way while lying down. Furthermore, touch input and the need to choose from several images/pictograms may not be the most adequate in more stressful situations (e.g., when the user is feeling unwell and wants to ask for help immediately). Other types of inputs, such as breathing (El-

[a] https://orcid.org/0000-0002-4094-7982
[b] https://orcid.org/0000-0001-6743-4530
[c] https://orcid.org/0000-0002-7675-1236

Rocha, A., Santos, D., Sánchez, F., Aguiar, G., Ramos, H., Ferreira, M., Bastos, T., Oliveira, I. and Teixeira, A.
Gesture Recognition for Communication Support in the Context of the Bedroom: Comparison of Two Wearable Solutions.
DOI: 10.5220/0011991700003476
In *Proceedings of the 9th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2023)*, pages 65-74
ISBN: 978-989-758-645-3; ISSN: 2184-4984

sahar et al., 2021; Wang et al., 2022) and brain signals (Peterson et al., 2020; Luo et al., 2022), present other limitations, including the need to wear intrusive sensors, being cumbersome to use especially while lying in bed, and/or requiring a considerable amount of initial training effort. There is thus a need to develop a solution for supporting communication that is adequate for use while in bed (both day and night), being as unobtrusive as possible, and that is also low-cost.

A solution based on the use of gestures can be a suitable option, since it does not require the direct physical use of a device, requiring only the user to carry out movements with a part of the body, such as the arm. In the context of human-human communication, a Personal Gesture Communication Assistant has been proposed, which recognizes gestures using a camera and machine learning (Ascari et al., 2019). However, no solution has been found that relies on gestures to enable remote two-way communication between a person with communication impairments and a caregiver, in the context of the in-bed scenario, besides previous work by our group (Guimaraes, 2021; Guimarães et al., 2021; Rocha et al., 2022; Santana, 2021; Santana et al., 2022).

Concerning gesture recognition, it typically relies on the data provided by one or more types of sensors, such as ambient (e.g., cameras, radars) or wearable (e.g., smartwatch). Although these sensors/devices are suitable for the target scenario, they usually have a non-negligible cost. The cameras have the additional disadvantage of commonly raising privacy issues for the users, even if RGB images are not used, especially in sensitive home divisions, such as the bedroom.

In the scope of the AAL APH-ALARM project[1], we previously carried out studies with both a smartwatch and a radar for gesture recognition while in bed. However, the radar work is still exploratory (Santana, 2021; Santana et al., 2022) and better results were achieved with the smartwatch (Guimarães et al., 2021; Guimaraes, 2021). Nonetheless, smartwatches have more features than what is needed for our proposal (only the accelerometer and gyroscope are used). Moreover, they have relatively large dimensions and weight, which make them uncomfortable to use, especially during prolonged use, including during sleep. Therefore, the main objective of this contribution is to improve on the proposed smartwatch-based communication support system, by exploring wearable alternatives to the smartwatch that are simpler and less expensive, with a better form factor, being thus more comfortable to use.

The use of an affordable and small size wearable has already been explored before (Zhao et al., 2019), but the authors did not compare it with a larger, more expensive wearable device. Some works compared different wearables, but they were placed at different body parts (e.g., finger and arm (Kurz et al., 2021)) or they included different sensor types (Kefer et al., 2017). Others compared only the sensors integrated in the same device (Le et al., 2019). To the best of our knowledge, there are no contributions comparing two types of wearable devices with the same sensors, used at the wrist.

To achieve our objective, we evaluated two alternatives regarding the wearable sensors used for gesture classification in our proposed system: a smartwatch (Oppo Watch) and a MbientLab module (MMR). This involved the evaluation of the performance of two models built with sensor data provided by each device, based on arm gesture data acquired from six different subjects while lying down.

[1] https://www.aph-alarm-project.com

2 SENSOR-BASED COMMUNICATION SUPPORT SOLUTION: OVERVIEW

For context, this section begins by giving an overview of the proposal for a communication support solution based on gestures that our group has been working on (Guimarães et al., 2021; Guimaraes, 2021; Rocha et al., 2022). Then, it presents the set of gestures selected for the current study, as well as the two variants of the solution that will be evaluated.

The target scenario of the system is the in-bed scenario, where a person with communication difficulties is lying in bed alone and may want to communicate with another person (e.g., caregiver, relative, friend) who is in a different division of the home, or even outside the home. The target primary users are people with communication difficulties, but without severe understanding difficulties and retaining motor function in at least one of the arms. The secondary users are the persons the primary user wants to communicate with.

2.1 Proposed Solution

An overview of the solution we propose is illustrated in Figure 1a. This solution relies on the use of arm gestures by the primary user to generate simple messages that are sent to the secondary user. To enable gesture input, the solution includes a corresponding modality (its main components are illustrated in Figure 1b), which relies on data sent by sensors worn by the primary user to decide on the gesture being per-

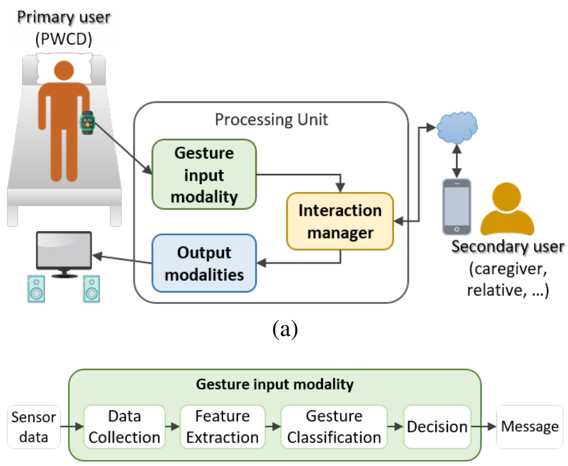


Figure 1: Overview of the proposed solution for communication support, including the users, sensors/devices, and interaction components (a), and main components of the gesture input modality (b).

formed and therefore the message to be sent.

The message constructed based on the gesture(s) is sent to the secondary user's smartphone through an Interaction Manager (IM), who is responsible for managing the exchange of information between the different modalities and applications of the system. After receiving a message, the secondary user can choose to send a message back, including pre-defined questions, to which the primary user can answer also using gestures. The information sent to the primary user is presented by one or more output modalities using a display and/or speakers in the bedroom.

2.2 Gesture Set

The selection of gestures for supporting communication took into account both the target scenario and the target primary users. Firstly, the gestures should be easy to execute while lying in bed and may rely on the ambient (in this case, the bed or mattress). Secondly, they should be easy to explain and understand, but also remember. Therefore, it should be possible to associate each of them with a specific meaning that makes sense in the considered context. To ensure the suitability of the gestures, the final choice resulted from the feedback obtained during discussion involving speech and language therapists with experience with people that have communication difficulties.

The set of arm gestures explored in this contribution is described in Table 1. These are the same gestures considered in previous work (Guimarães et al., 2021; Guimaraes, 2021), with the exception of one (clockwise circle with the hand in the air), which was excluded since we considered it can be more difficult to execute by some users (e.g., older adults).

2.3 Two Variants

We implemented two variants of the solution described above, by implementing two different versions of the gesture input modality. This modality includes a data acquisition, feature extraction, gesture classification, and decision modules (Figure 1). Gesture classification relies on a model that recognizes the current gesture, which is trained using machine learning and features extracted from sensor data acquired from a set of subjects.

The two variants use two different wearable devices: a smartwatch and a simpler, more affordable sensor. More concretely, the devices were a Wear OS smartwatch, namely a Oppo Watch, and a module from MbientLab, more specifically the MMR module. Both include a 3-D accelerometer, gyroscope, and magnetometer.

The Oppo Watch had a launch price of over 200 dollars/euros, in 2020. Its body dimensions (height x width x depth) are 46 x 39 x 13 mm or 42 x 37 x 13 mm (both with heart rate monitor), and it weighs 40 or 30 g, respectively (Oppo, 2021). On the other hand, the MMR module was selling for 80 dollars (MbientLab, Inc., 2022), and its dimensions (without case) are 29 x 18 x 6 mm and its weight is 6 g.

Concerning the mentioned sensors, although the

Table 1: Set of gestures considered for the proposed system, including the description and also the possible meaning of each gesture.

Gesture	Description	Meaning
Twist	Twist the wrist from left to right and vice-versa, preferably with an angle between 0 an 45 degrees between the bed and forearm.	Ask for immediate help (e.g., feeling unwell)
Come (to me)	Move the forearm towards the arm until a 45-degree angle is formed between the arm and forearm, as if calling for someone.	Ask for not so urgent help (e.g., help getting up from bed)
Knock	Knock with the hand on the mattress, while keeping the arm close to the body.	Affirmative meaning (yes), when answering a question
Clean	Move the hand and arm from left to right and vice-versa, with the hand in contact with the mattress.	Negative meaning (no), when answering a question

differences among different sensors of the same type may not be as considerable as they used to be in the past, there are still some disparities for the two studied devices, as can be seen in Figure 2. Although some differences can be due to the placement of the sensors, we can see that the the output of the smartwatch (SW) shows a more clear pattern for the accelerometer's z-axis and the gyroscope's y-axis, when compared with the MbientLab module (MB).

3 COMPARATIVE EVALUATION OF THE TWO VARIANTS

Evaluation focused on the gesture recognition performance of the two variants of the proposed solution. Sensor data provided by the devices used in the two variants were acquired from six volunteers, all male students at the Department of Electronics, Telecommunication, and Informatics of the University of Aveiro. They were all right-handed and 20 or 21 years old (mean $\pm$ standard deviation of 21.7 $\pm$ 0.5). All participants read and signed an informed consent.

3.1 Setup and Protocol

For each participant, the wearable devices were attached to the wrist of the dominant arm, next to each other, as shown in Figure 3. The MbientLab device was placed inside a case and attached to the wrist using a Velcro strap. Both devices were always placed in a similar way for all participants, to ensure that the orientation of their coordinate systems was similar among devices and participants.

Figure 3: Setup used for the wearable devices attached to the wrist.

The experiment was carried out at a room of our institute, where a "bed" was set up using two tables

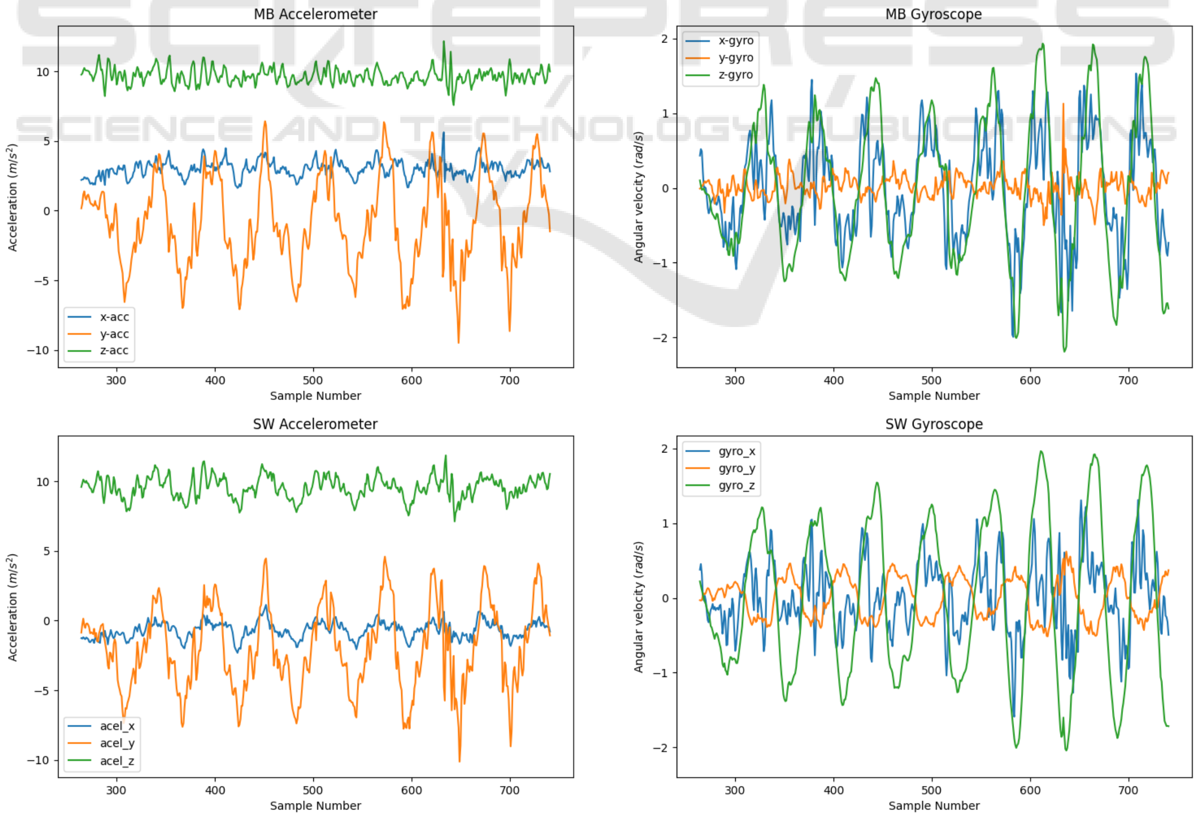


Figure 2: Example of the 3-D signals for the accelerometer and gyroscope of both explored devices (smartwatch – SW, and MbientLab module – MB), for several repetition of a given gesture by a given subject.

and six cushions from sofas tied together using ribbons as the mattress. A bench was also placed next to the "bed" to simulate the floor, having a height between them that is similar to a typical height between a bed and the floor.

Since the main objective is to compare the results obtained with the two devices, it was also necessary to ensure that their signals were synchronized. To achieve this, each performed data recording included a single execution of specific movement. For simplicity, we used one of the defined arm gestures, in this case, the "come" gesture.

For each subject and arm gesture included in Table 1, the following experimental protocol was carried out by the participant: (1) begin by lying still; (2) after the recording is started, perform the synchronizing gesture a single time; (3) lie still for around 3 seconds; (4) perform the indicated gesture repeatedly until the end of the recording.

Since other movements are expected to be carried out by a person in the bed scenario, other gestures/activities were additionally included in the protocol, namely lying down without moving, rotating the body to the left and/or right, lying to standing up next to bed, standing next to bed to lying, and any other movement while lying down (chosen by the participant).

A sequence with the set of gestures and other movements was performed and recorded twice per participant. The order of execution was randomly chosen for each participant and sequence.

3.2 Data Acquisition and Pre-Processing

For both devices, the data were acquired at an approximate rate of 50 Hz for all sensors. The only exception was the MbientLab module's magnetometer, for which a rate of 25 Hz was selected, since 50 Hz was not available.

For each recording, the signals from the two devices were synchronized based on the cross-correlation between the signals corresponding to the sum of the gyroscope values for the three axes. After synchronization, the original signals were segmented by selecting only the data corresponding to the gesture repetitions.

This segmentation was carried out automatically based on the MbientLab module's gyroscope signals. Firstly, the gesture used for synchronization was ignored by performing the following procedure for each axis: (1) find the first frame for which the absolute value is higher than 2 rad/s; (2) starting at that frame, count the number of frames for which the absolute value is lower than 0.7 rad/s; and (3) find the frame for which the previous count is higher than 45 frames. All data before this last frame were discarded. For the remaining data, and for each axis, the beginning of the gesture repetition was defined by finding the first frame for which the absolute value is higher than 0.5 rad/s. The ending of the gesture repetition interval was also found using the same threshold, but starting at the end and going backwards. Finally, the signals were segmented based on the minimum initial frame number and the maximum final frame number, when taking into account the three axes. All threshold values were chosen empirically.

Successful synchronization and segmentation was verified based on visual observation for all recordings. Besides synchronization/segmentation, no further processing of the raw signals, such as filtering, was carried out.

Several time-domain features were then computed over the accelerometer and gyroscope signals only. The magnetometer was not used, since visual observation of the raw signals from the two different devices for the same recording showed considerable differences between them, most likely due to the lack of proper calibration for one or both devices.

The features, which are listed and described in Table 2, were selected based on previous work with the smartwatch (Guimaraes, 2021) and expanded with the following: range without outliers, interquartile range (IQR), root mean square (RMS), and median absolute deviation (MAD). All features were computed for each sensor and axis, except for the Pearson's correlation coefficient, which was computed for each sensor and axes pair (xy, yz, and xz), as well as considering all axes combinations between the two sensors. This resulted in 93 features.

Since some of the defined arm gestures (e.g., "clean" and "knock") have more movement on a given 2D plane, similar features were computed for the signals corresponding to the magnitude of the vector on each plane, using (1) for xy-plane and a similar equation for the other two planes. In (1), s is the 2D vector, and s_x and s_y correspond the value of s in the x- and y-axis, respectively.

$$\|s\| = \sqrt{s_x^2 + s_y^2} \tag{1}$$

The features were extracted considering a window of 2 s. This duration was chosen based on our previous results with the smartwatch (Guimaraes, 2021). Since the evaluation will consider the recognition result for each window separately, an overlap of 99% between consecutive windows was used for the evaluation to obtain the largest number of examples possible.

Table 2: Features extracted from the sensor data. All features were computed for each sensor and axis/plane, with the exception of the Pearson's correlation coefficient, which was calculated for each pair of axes/planes within the same sensor and also between the two sensors.

Name	Description
Mean	Mean considering all samples
Median	Median considering all samples
Mean-median difference	Difference between the mean and median value
Standard Deviation	Standard deviation considering all samples
Variance	Variance considering all samples
Range	Difference between maximum and minimum values of the signal
IQR	Interquartile range, i.e., the difference between the third quartile (Q3) and the first quartile (Q1) considering all samples
Range without outliers	Range excluding outliers, where an outlier is any value above Q3 or below Q1.
RMS	Root mean square considering all samples
MAD	Median absolute deviation, i.e., the median of the absolute differences between each sample and the median
Skewness	Measure of asymmetry of the probability distribution of the signal about its mean
Kurtosis	Measure of the "tailedness" of the probability distribution of the signal
Integral	Area under the curve
Correlation	Pearson's correlation coefficient

Finally, the features were scaled using the "RobustScaler" provided by Python's "scikit-learn" package (Pedregosa et al., 2011).

3.3 Datasets

The resulting dataset used for evaluation was similar for both wearable devices. Each dataset includes 186 features. The class corresponds to the name of the gesture performed by the user, with the other movements being considered as a single class named "other".

Since the datasets were not balanced in terms of classes (gestures) and groups (subjects), for each dataset, the same number of examples was randomly selected per gesture and participant, with that number corresponding to the minimum number of examples when considering each gesture/participant combination. For each device, the resulting balanced dataset included 15,600 examples (2,600 per subject and 3,120 per gesture).

3.4 Classifier

The classifier selection was also based on our previous results (Guimaraes, 2021). The best classifier for the subject independent solution was the support vector machines (SVM) algorithm, with a linear kernel. For subject-dependent, the best algorithm was the random forest followed by the SVM, but both led to a similar performance. Therefore, in the current work, the SVM is used in both cases.

The model evaluation was performed using Python's "scikit-learn" package (Pedregosa et al., 2011). The default values for the SVM's hyperparameters were used, except for the kernel, which was the linear kernel.

3.5 Evaluation Approach

Besides comparing the two devices, we also wanted to compare two different solutions – subject dependent and subject independent – to investigate if it is possible to train a single model that can be used with never-seen users (subject independent), or if it is necessary to train a model for each new user (subject dependent).

For the subject dependent case, we applied the stratified 10-fold cross-validation approach to the data of each subject separately. This approach consists of dividing the considered dataset into 10 sub-samples with the same size, in a random way, but ensuring that the number of examples from each class are approximately the same for each sub-sample. Then, 10 iterations are performed. For each iteration, one of the sub-samples is used for testing, while the remaining 9 sub-samples are used for training the model. The test set is always different for each iteration.

As the subject independent solution consists of using data from a group of subjects to train the model and then using the resulting model to classify examples from new subjects, in this case, we applied the leave-one-subject-out cross-validation (LOSO-CV) approach to the whole dataset. This approach involves as many iterations as the total number of subjects in the dataset. In each iteration, the data corresponding to all subjects except one is used for training, while the data of the remaining subject is used for testing. A different subject is used as the test set in each iteration. To obtain more than one result per subject, for each LOSO-CV iteration, we further used an adapted stratified 10-fold CV approach, where 10 iterations are further carried out. For each inner iteration, 9 sub-samples from the training set are used to train the model, while 1 sub-sample from the test set is used for testing. The used sub-samples are always different for each inner iteration.

In both subject dependent and independent cases, the following metrics were computed: overall accuracy; class F1 score, i.e., the F1 score for each considered class or gesture type; and overall F1 score (mean of all class F1 scores). The false positive rate (FPR) considering all arm gestures as the positive class and "other" as the negative class was also computed.

4 RESULTS

This section presents the results obtained for the subject dependent and independent solutions.

4.1 Subject Dependent

For the subject dependent solution, the model classified all examples correctly, for all subjects and CV iterations, and for both explored devices.

These results are in line with the results obtained previously by our group for the smartwatch (Guimaraes, 2021). Furthermore, they show that, in this case, not only both devices can be used interchangeably, but are also suitable for recognizing gestures in bed. Nevertheless, it is important to note that the considered gesture set is relatively small, and the number of different "other" gestures/activities is limited. Moreover, the number of participants is also small and represent a very specific subject group.

Although the subject dependent solution has a very good performance, it has the disadvantage of requiring the collection of data from each new subject before they can start using the system.

4.2 Subject Independent

The subject independent solution does not have that disadvantage. On the other hand, it is more challenging to achieve a good performance, since it is usually more difficult to classify gestures for a never-seen subject using a model trained with data from a group of other subjects. The results obtained for this case are reported and discussed below.

4.2.1 Overall

Table 3 shows the overall results achieved for each device, considering all subjects and gestures, as well as for the difference between them. The results include the mean, standard deviation, median, minimum, and maximum values. As the obtained results are very similar for both wearables, to better compare them, the device differences are presented in Figure 4.

Table 3: Mean, standard deviation (Std), mean, minimum (Min), and maximum (Max) values achieved with the smartwatch (SW) and MbientLab module (MB), when considering all subjects and gestures, for each considered metric (accuracy, F1 score, and false positive rate – FPR), in the case of the subject independent solution. The results for the device differences are also included (SW-MB).

Metric	Statistic	SW	MB	SW-MB
Accuracy (%)	**Mean**	98.4	98.4	0.0
	Std	1.7	1.8	0.8
	Median	98.8	99.4	0.0
	Min	93.8	94.2	-2.7
	Max	100.0	100.0	1.5
F1 score (%)	**Mean**	98.4	98.4	0.0
	Std	1.7	1.8	0.8
	Median	98.8	99.4	0.0
	Min	93.9	94.3	-2.8
	Max	100.0	100.0	1.5
FPR (%)	**Mean**	2.0	1.5	0.5
	Std	3.7	3.0	3.3
	Median	0.0	0.0	0.0
	Min	0.0	0.0	-5.8
	Max	15.4	11.5	13.5

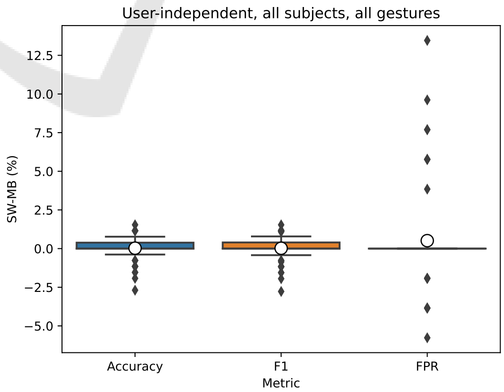


Figure 4: Boxplot of the device differences for the accuracy, F1 score, and false positive rate (FPR), considering all subject and gestures. MB and SW stand for MbientLab module and smartwatch, respectively. The circle indicates the mean value.

The performance for each device is worse comparing with the subject-dependent case, as expected. However, it is only slightly worse, with mean and median values of 98% for accuracy and F1 score, and ≤2% for FPR, and low standard deviation, for both devices.

In addition, the results are better than those we achieved previously with the smartwatch (Guimaraes, 2021), but it is necessary to consider that the current work explored 4 instead of 5 arm gestures, and a much larger number of features was used.

From Figure 4, we can see that there are some cases where one of the devices outperforms the other. However, for accuracy and F1 score, the absolute difference is always lower than 3%. For FPR, the device difference values vary more, ranging between -6% and 14%. Nevertheless, the values different from 0% are all outliers.

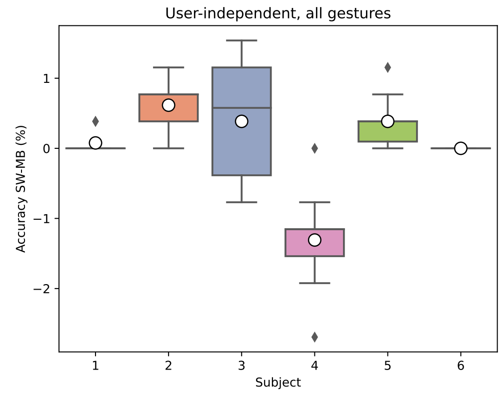
To verify if the differences between devices are statistically significant, we carried out the Wilcoxon signed-rank test over the device differences, separately for each considered metric. A non-parametric test was chosen since the result of the Shapiro-Wilk test does not indicate that the data have a normal distribution (p-value<0.05). Both tests were performed using Python's "scipy" package. The Wilcoxon test results showed that the device differences are not significant (p-value≥0.05).

4.2.2 Subject Effect

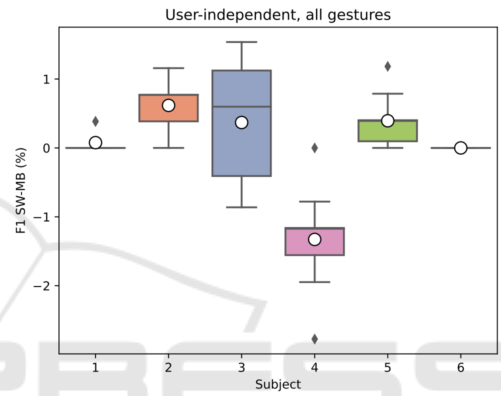
To better understand if this is also true when analyzing each subject individually, we further investigated the results per participant, considering all gestures, which are shown in Figure 5.

We can see that there is a variation between the different subjects, with some presenting a higher device difference variability than others. For the accuracy and F1 score, Participant 3 has the highest variability, but with both positive and negative values. For Participants 2 and 5, the differences are always ≥0%, while the opposite happens for Participant 4 (≤0%). The other two participants show very small or no variability (median and mean of 0.0% or 0.1%). When considering the FPR metric, variability is very small for all participants, except Participant 2 (best overall performance for SW) and Participant 4 (best overall performance for MB).

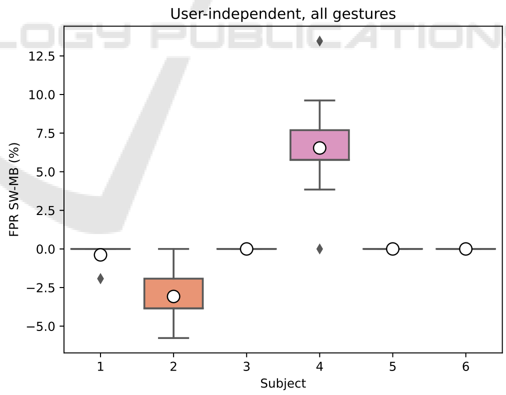
The results of the Wilcoxon signed-rank test for each subject show that, for accuracy and F1 score, there is no significant difference between the two devices for Participant 1, 3 and 6 (p-value≥0.05). On the other hand, the difference is significant for Participants 2, 4 and 5 (p-value<0.05). For FPR, there is a significant difference only for Participants 2 and 4. Therefore, considering all three metrics, device performance is similar for half of the participants, while one of the devices outperforms the other for half of the participants. The SW performs better than the MB for two participants, and the opposite is observed for one participant. Nevertheless, it is worth noting that the maximum absolute difference for accuracy and F1 score is always lower than 3%.



(a)



(b)



(c)

Figure 5: Boxplot of the device difference results for the (a) accuracy, (b) F1 score, and (c) false positive rate (FPR), for each subject, considering all gestures. MB and SW stand for MbientLab module and smartwatch, respectively. The circle indicates the mean value.

4.2.3 Gesture Effect

Focusing on the different gesture types, the F1 score results per gesture, presented in Figure 6, show that the performance is the same for both devices in most cases (apart from some outliers) for all gestures, except “other”. For “other”, variability is higher when excluding the outliers (-1.5 to 2.5%), but both the mean and median are of 0%.

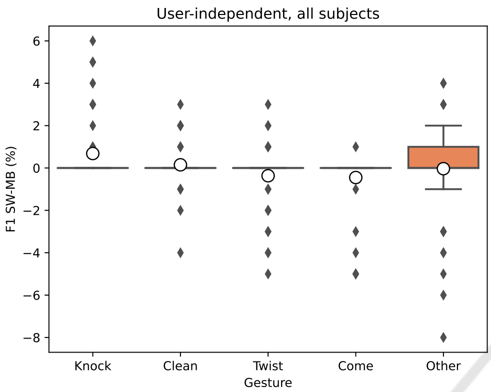


Figure 6: Boxplot of the device difference results for the F1 score of each gesture, considering all subjects. MB and SW stand for MbientLab module and smartwatch, respectively. The circle indicates the mean value.

Nevertheless, there is some asymmetry in the results for some gestures, with the results of the Wilcoxon signed-rank test for each gesture showing that there is a significant difference (p-value<0.05) between devices for “knock”, “twist”, and “come”. In the case of “knock”, the best device is SW, while the best performance for “twist” and “come” is achieved with the MB device. However, despite a statistically significant difference, it is still relatively small, with a maximum absolute difference value of 8%.

5 CONCLUSION

The main aim of this work was to evaluate two wearable alternatives, with different dimensions and costs, for recognizing gestures to be used in a communication support solution. The context of this study was the bedroom scenario, where a person with communication difficulties is lying in bed alone and uses gestures to communicate remotely with another person, such as a caregiver, family member, or friend.

Accelerometer and gyroscope data provided by two different devices (MbientLab module and smartwatch) were collected from six subjects while they performed four relevant arm gestures, as well as other gestures or movements. Several features were extracted from the synchronized signals. For each wearable, we then evaluated a model built using those features and the support vector machines algorithm.

For a subject dependent solution, the model was able to correctly classify all examples, using any of the two devices. For the subject independent case, which is more challenging, the performance was still quite good for both wearables (mean accuracy and F1 score of 98%, and false positive rate of 2%). When considering each subject and each gesture, there were some cases where one of the devices outperformed the other, but the best device varied. When taking into account all subjects and gestures, there was no significant differences between the smartwatch and the MbientLab module.

We can conclude that, overall, both alternatives provided promising and similar results, being possible to rely on a simpler wearable with lower cost and smaller size for recognizing the arm gestures considered for supporting communication in the bedroom scenario.

6 FUTURE WORK

A limitation of this study is that the participants are healthy young adults. Although communication difficulties can result from different problems and thus affect people of different ages, the volunteers of this study do not represent for example older people, who may also have slower or more restricted movements as a consequence of aging. Another limitation is the fact that data were acquired while the subjects were are lying in a specific posture, i.e., lying on their back, and only one of the arms was used for gesture execution.

In the future, the results of this study should be confirmed with data from a greater number of subjects with more varied ages, including older subjects. It would also be important to explore other postures in bed (e.g., lying on the side) and outside the bed (e.g., sitting on a sofa), and both arms for executing the gestures. A greater variety of “other” gestures/activities should also be included in the dataset. Since the number of features used in this study was relatively high, another interesting aspect to investigate is if similar results can be obtained with a smaller set of features.

ACKNOWLEDGMENTS

The authors wish to thank the volunteers who participated in the experiment, and Prof. José Maria Fernandes for providing the MbientLab device. This work was supported by EU and national funds through the Portuguese Foundation for Science and Technology (FCT), in the context of the AAL project APH-ALARM (AAL/0006/2019) and funding to the research unit IEETA (UIDB/00127/2020).

REFERENCES

Allen, M., McGrenere, J., and Purves, B. (2007). The design and field evaluation of phototalk: A digital image communication application for people. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '07, page 187–194, New York, NY, USA. Association for Computing Machinery.

Ascari, R. E. O. S., Silva, L., and Pereira, R. (2019). Personalized interactive gesture recognition assistive technology. In *Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems*, IHC '19, New York, NY, USA. Association for Computing Machinery.

Elsahar, Y., Hu, S., Bouazza-Marouf, K., Kerr, D., and Mansor, A. (2019). Augmentative and alternative communication (aac) advances: A review of configurations for individuals with a speech disability. *Sensors*, 19(8).

Elsahar, Y., Hu, S., Bouazza-Marouf, K., Kerr, D., Wade, W., Hewett, P., Gaur, A., and Kaushik, V. (2021). A study of decodable breathing patterns for augmentative and alternative communication. *Biomedical Signal Processing and Control*, 65:102303.

Guimaraes, A. M. (2021). Bedfeeling - sensing technologies for assistive communication in bed scenarios. Master's thesis, University of Aveiro, Aveiro, Portugal.

Guimarães, A., Rocha, A. P., Santana, L., Oliveira, I. C., Fernandes, J. M., Silva, S., and Teixeira, A. (2021). Enhanced communication support for aphasia using gesture recognition: The bedroom scenario. In *2021 IEEE International Smart Cities Conference (ISC2)*, pages 1–4.

Kane, S. K., Linam-Church, B., Althoff, K., and McCall, D. (2012). What we talk about: Designing a context-aware communication tool for people with aphasia. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '12, page 49–56, New York, NY, USA. Association for Computing Machinery.

Kefer, K., Holzmann, C., and Findling, R. D. (2017). Evaluating the placement of arm-worn devices for recognizing variations of dynamic hand gestures. *J Mob Multimed*, 12(3–4):225–242.

Kurz, M., Gstoettner, R., and Sonnleitner, E. (2021). Smart rings vs. smartwatches: Utilizing motion sensors for gesture recognition. *Appl Sci*, 11(5).

Laxmidas, K., Avra, C., Wilcoxen, C., Wallace, M., Spivey, R., Ray, S., Polsley, S., Kohli, P., Thompson, J., and Hammond, T. (2021). Commbo: Modernizing augmentative and alternative communication. *International Journal of Human-Computer Studies*, 145:102519.

Le, T.-H., Tran, T.-H., and Pham, C. (2019). The internet-of-things based hand gestures using wearable sensors for human machine interaction. In *2019 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, pages 1–6.

Love, T. and Brumm, K. (2012). CHAPTER 10 - Language processing disorders. In Peach, R. K. and Shapiro, L. P., editors, *Cognition and Acquired Language Disorders*, pages 202–226. Mosby, Saint Louis.

Luo, S., Rabbani, Q., and Crone, N. E. (2022). Brain-computer interface: Applications to speech decoding and synthesis to augment communication. *Neurotherapeutics*, 19(1):263–273.

MbientLab, Inc. (2022). MMR – MetaMotionR. Accessed: 6 Nov. 2022.

Obiorah, M. G., Piper, A. M. M., and Horn, M. (2021). Designing aacs for people with aphasia dining in restaurants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA. Association for Computing Machinery.

Oppo (2021). Oppo Watch Specification. Accessed: 8 Dec. 2022.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Peterson, V., Galván, C., Hernández, H., and Spies, R. (2020). A feasibility study of a complete low-cost consumer-grade brain-computer interface system. *Heliyon*, 6(3):e03425.

Rocha, A. P., Guimarães, A., Oliveira, I. C., Nunes, F., Fernandes, J. M., Oliveira e Silva, M., Silva, S., and Teixeira, A. (2022). Toward supporting communication for people with aphasia: The in-bed scenario. In *Adjunct Publication of the 24th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '22, New York, NY, USA. Association for Computing Machinery.

Santana, L., Rocha, A. P., Guimarães, A., Oliveira, I. C., Fernandes, J. M., Silva, S., and Teixeira, A. (2022). Radar-based gesture recognition towards supporting communication in aphasia: The bedroom scenario. In Hara, T. and Yamaguchi, H., editors, *Mobile and Ubiquitous Systems: Computing, Networking and Services*, pages 500–506, Cham. Springer International Publishing.

Santana, L. F. d. V. (2021). Exploring radar sensing for gesture recognition. Master's thesis, University of Aveiro, Aveiro, Portugal.

Wang, Q., Zhao, J., Xu, S., Zhang, K., Li, D., Bai, R., and Alenezi, F. (2022). ExHIBit: Breath-based augmentative and alternative communication solution using commercial RFID devices. *Information Sciences*, 608:28–46.

Zhao, H., Wang, S., Zhou, G., and Zhang, D. (2019). Ultigesture: A wristband-based platform for continuous gesture control in healthcare. *Smart Health*, 11:45–65.