applied sciences

MDPI

Article

Person Independent Recognition of Head Gestures from Parametrised and Raw Signals Recorded from Inertial Measurement Unit

Anna Borowska-Terka * and Pawel Strumillo

Faculty of Electrical, Electronic, Computer and Control Engineering, Institute of Electronics, Lodz University of Technology, 211/215 Wolczanska Str., 90-924 Lodz, Poland; pawel.strumillo@p.lodz.pl

* Correspondence: anna.borowska-terka@p.lodz.pl

Received: 10 May 2020; Accepted: 17 June 2020; Published: 19 June 2020

Abstract: Numerous applications of human–machine interfaces, e.g., dedicated to persons with disabilities, require contactless handling of devices or systems. The purpose of this research is to develop a hands-free head-gesture-controlled interface that can support persons with disabilities to communicate with other people and devices, e.g., the paralyzed to signal messages or the visually impaired to handle travel aids. The hardware of the interface consists of a small stereovision rig with a built-in inertial measurement unit (IMU). The device is to be positioned on a user's forehead. Two approaches to recognize head movements were considered. In the first approach, for various time window sizes of the signals recorded from a three-axis accelerometer and a three-axis gyroscope, statistical parameters were calculated such as: average, minimum and maximum amplitude, standard deviation, kurtosis, correlation coefficient, and signal energy. For the second approach, the focus was put onto direct analysis of signal samples recorded from the IMU. In both approaches, the accuracies of 16 different data classifiers for distinguishing the head movements: pitch, roll, yaw, and immobility were evaluated. The recordings of head gestures were collected from 65 individuals. The best results for the testing data were obtained for the non-parametric approach, i.e., direct classification of unprocessed samples of IMU signals for Support Vector Machine (SVM) classifier (95% correct recognitions). Slightly worse results, in this approach, were obtained for the random forests classifier (93%). The achieved high recognition rates of the head gestures suggest that a person with physical or sensory disability can efficiently communicate with other people or manage applications using simple head gesture sequences.

Keywords: electronic human-machine interface; blindness; gesture recognition; inertial sensors; IMU

1. Introduction

Human–System Interaction (HSI) is currently actively pursued as a separate research field dedicated to the development of new technologies facilitating human communication with systems [1]. Depending on the application, such interaction systems are also referred to as Human–Computer Interfaces (HCI) or more generally human–machine interfaces. The design and construction of such systems requires an interdisciplinary research approach and involves knowledge of sensory perception mechanisms in humans, cognitive processes, and information processing, as well as basics of ergonomics. A well-designed user interface often determines the usefulness of an entire system [2].

Designing interfaces accessible for persons with sensory and motor disabilities is a particularly challenging research issue [3,4]. It is necessary to develop innovative solutions that enable handicapped users to communicate with such devices or systems. The interfaces use alternative, often multimodal communication methods that compensate for the diminished or lost sensorial or physical functions [5].

Individuals with hearing loss use sign language (one of the so-called visual-spatial languages) and are aided by sign language computer applications. For people with physical disabilities, special input-output devices utilizing inertial sensors and innovative interfaces are built so that "contactless" communication with a computer via, among others, Brain–Computer Interfaces (BCI) [6] and video interfaces, which are capable of tracking eye movements [7] or detecting intentional blinks [8], have been developed. The work reported in [9] shows that people with cerebral palsy can communicate with the use of personalized gesture datasets while aided by an intelligent user interface. The visually impaired, on the other hand, interact with computers using speech synthesizers and the so-called Braille displays that allow them to type in and display the Braille alphabet. A more complex problem for a blind person is handling devices while in motion. The basic mobility aid that the blind use is a white cane, which engages one hand and hence, complicates the operation of an additional device, e.g., a navigation tool during a walk.

The control of computer applications is traditionally performed using a keyboard and/or a computer mouse or a touch screen. However, solutions that enable people with serious physical disabilities to work with computers are increasingly more common. In [10], an interface based on the recognition of head movements and simultaneous lip movements designed for users with limb disability was reported. It is, however, not only computer applications that can be controlled in a non-contact manner. It was shown that the movement of an electric wheelchair of a disabled person can be controlled by means of head movements [11–13] or eye gaze [14] and even EEG signals [15].

Recent advances in electromechanical MEMS (Micro Electro Mechanical Systems) technologies have made it possible to develop miniature and cheap inertial sensors, and consequently use them in human–computer interfaces [16]. In [17], in order to identify eight different types of physical activity, a three-axis accelerometer placed on the wrist of the dominant hand was used. In [18], five acceleration sensors were used to recognize the movement of hands and to distinguish gait from immobility. The sensors were placed on the chest and on all the limbs. Similarly to our work, 3-axis compass, 3-axis accelerometer and 3-axis gyroscope were mounted on the user's head and used to identify six different head gestures [19]. In [20], a system that can be utilized to monitor the body movements of people with neurodegenerative diseases was reported. In this system, signals recorded from three sensors were analyzed, one of them was mounted on the head and the other two on the shins. The fusion of signals acquired from an accelerometer and surface EMG electrodes was used in [21] to assess Parkinson's disease-related symptoms. In [22], simultaneous processing of images and signals from inertial sensors was employed to estimate six degree-of-freedom head movements. In the above-mentioned studies, the first processing step in analyzing IMU signals is to parametrize the signals for predefined analysis time windows [17,18]. The extracted parameters are then used to build appropriate classifiers that recognize given types of movements.

In a recent work [23], novel methods for recognizing human physical activity that are based on the so called symbolic representation algorithms were presented. Using three database sets, the authors have shown that their approach performs best in terms of accuracy, processing time, and memory consumption in comparison to other classic approaches that were based on supervised classification techniques. In another very recent work [23], IMU were applied to measure the 3D range of motion of the trunk and lower limb joints. The results of this study show that inertial sensors can be successfully applied to investigate maladaptive movement strategies.

In our study, we propose a hands-free interface enabling blind persons to control the menu of a navigation device by means of head movements. We also envision that such an interface can serve as a communication aid for people with serious motor disabilities, e.g., for the paraplegic individuals.

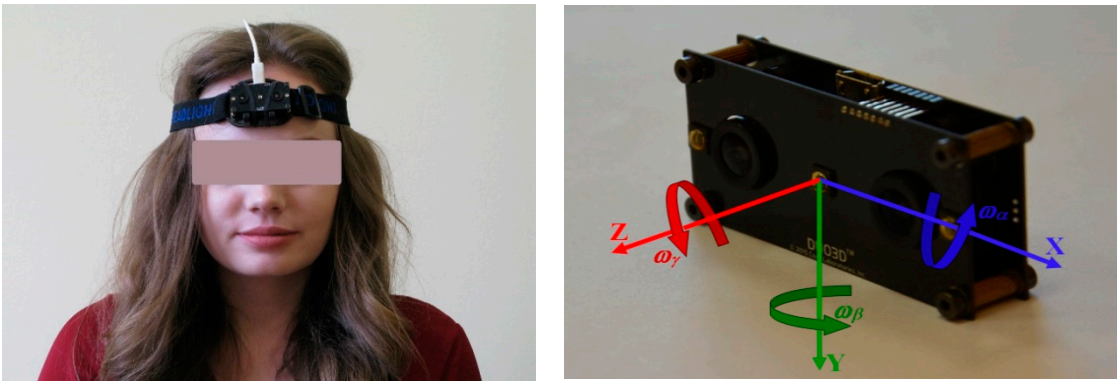
The main contribution of our work, which has not been addressed in earlier studies [17–19,24], is the comparison of the parametric and the time domain representations of IMU signals on which different classifiers are trained to recognize four head movement patterns. Similarly to the work reported in [25], in which IMU were used to recognize types of physical activities, we studied the dependence of different lengths of time windows on the parametric representation of inertial signals

to recognize different types of head gestures. In our work, we show that time domain approaches to analysis of inertial signals can outperform parameter-based ones and are less computationally demanding. The latter feature is particularly important for mobile implementations of the interface. The ultimate motivation for our research is to develop user-friendly and efficient electronic travel aids that will help the visually impaired retain orientation and mobility in unfamiliar environments [26,27].

The remainder of this paper is organized as follows. In Section 2, we describe the experimental methods that we used for data recording and data pre-processing. We also shortly review the applied data classifiers. In Section 3, we present the head gesture recognition results for the parametric and time domain representations of IMU signals. Finally, in Section 4 we apprise the achieved results, show the envisioned applications of our study, and point out limitations of the presented work.

2. Materials and Methods

In our proof of concept approach, we have applied a DUO MLX device (see Figure 1b) equipped with a stereovision camera, a three-axis electronic gyroscope, three-axis acceleration sensor, a thermometer, and a magnetometer [28]. The device is of small form factor: 52 × 25 × 13 mm and weighs 12.5 g. The device is mounted on the user's forehead (see Figure 1a). The study was conducted with 65 persons. The participants were mainly second year university students, 44 of whom were women. The trials were approved by the bioethics commission of the Medical University of Lodz (No. RNN/261/16/KE). All the trial participants were informed about the purpose of the study, the materials used and their role in the trial sessions. After fixing the DUO MLX device on their forehead, the users were asked to perform rehearsal head movements and were instructed to proceed with the movements just within a comfort zone of their head positions. During collection of the data, the participants remained in a sitting position. Each participant was sitting straight in a chair and did not change position during the experiment and performed only the given motions (yaw, pitch, roll, and immobility). Each user had the DUO MLX device mounted rigidly on their forehead. The participant did not touch the DUO MLX device or change its position on the forehead during the experiment.



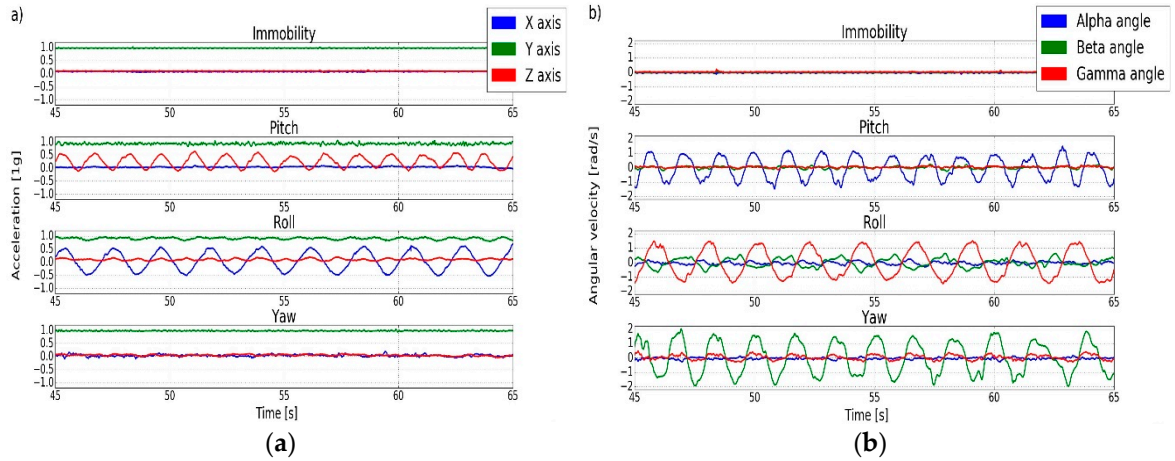
(**a**) (**b**)

Figure 1. (**a**) Placement of the inertial measurement unit on the user's head; (**b**) DUO MLX Stereo System equipped with inertial sensors with marked coordinate axes *Ox*, *Oy*, and *Oz*, and angular velocities ω_α, ω_β, and ω_γ.

The aim of the study was to automatically classify three basic head gestures: *roll* (head movement "sideways"), *pitch* (head movement "up-down") and *yaw* (head movement "left-right"), and *immobility* (head at rest). A single movement cycle lasted 2.5 s on average with 1.2 s standard deviation of the movement period. The recordings of test and learning datasets from one individual took approx. 4 min each.

Sample plots of the signals recorded by the acceleration sensor and the gyroscope are shown in Figure 2. The recorded accelerations a_x, a_y, and a_z along axes *Ox*, *Oy*, and *Oz* respectively are plotted in

Figure 2a and angular velocities ω_α, ω_β, and ω_γ (as shown in Figure 1b) recorded from the gyroscope are plotted in Figure 2b. All the signals are recorded at a sampling rate of 100 Hz.



(**a**) (**b**)

Figure 2. Example signal waveforms recorded by: (**a**) acceleration sensors and (**b**) gyroscope sensors.

2.1. Signal Recordings and Pre-Processing

Signal recordings from 65 participants were grouped into the following datasets:

1. The learning dataset recorded from a randomly selected 53, i.e., approx. 80% of all the 65 trial participants. Signals that were used as the learning set were recorded as follows: *10 s immobility—2.5 min of a specific motion type—10 s immobility*, the type of gestures performed by test participants were: yaw, roll, pitch, and immobility. The collection of the learning signals therefore consisted of 53 × 4 = 212 recordings.
2. Two types of testing datasets that have not been used for training the classifiers:

   - *T*_53_*set* that was recorded for 53 trial participants for whom the learning data were recorded,
   - *T*_12_*set* that was recorded for the rest of the trial participants, i.e., 12 individuals who did not take part in the recordings of the learning datasets,

   further, for each of the two testing data sets there were two different testing scenarios applied:

   - testing scenario *T*1 for which the dataset was recorded following the sequence of: *20 s of immobility—40 s of yaw head movement—20 s of immobility—40 s of roll—20 s of immobility—40 s of pitch—20 s of immobility*; thus, two types of the testing datasets were recorded *T*1_53_*set* and *T*1_12_*set*, respectively
   - testing scenario *T*2 for which the datasets were recorded according to the following procedure: head gesture sequences of *roll*, *yaw* and *pitch* gestures in random order for time periods lasting from 5 to 20 s and no immobility time gaps between the gestures; thus, two other types of the testing datasets were recorded: *T*2_53_*set* and *T*2_12_*set* respectively.

The signals recorded from the IMU were processed according to the two following procedures:

1. *With signal parameterization* (*SP*): for each of the six recorded IMU signals, i.e., three signals from the accelerometer (a_x, a_y, a_z) and three signals from the gyroscope (ω_α, ω_β, ω_γ), the following features were extracted: (1) average, (2) minimum, (3) maximum, (4) standard deviation, (5) kurtosis, (6) correlation coefficients for pairs of signals from the accelerometer and pairs of signals from the gyroscope, and (7) signal energy. Thus, the total number of parameters derived from the signals was: 6 signals × 7 parameters = 42 parameters.

These parameters were calculated for different time window widths, i.e., $T \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 1.0, 2.0\}$ given in seconds. The work focuses on the causal recognition mode of the gestures for which the time windows are determined for current and past samples only. The consecutive windows were shifted by a 0.1 s time step as shown in Figure 3.

2. *Time domain representation* (*TDR*): current samples of signals recorded from the accelerometer (a_x, a_y, a_z) and gyroscope (ω_α, ω_β, $\omega_{\gamma)}$ were used directly as six-element training vectors for the classifiers.

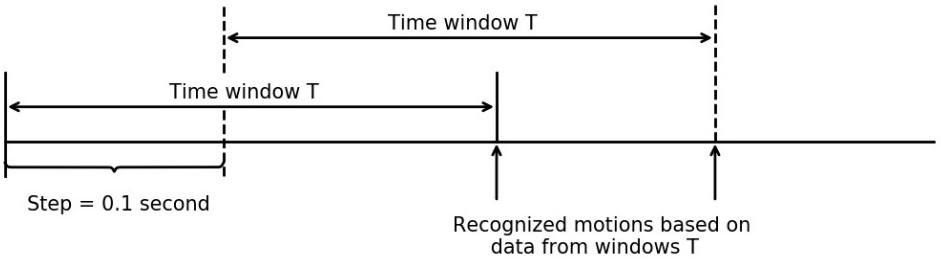


Figure 3. An example of two consecutive time windows for which signal parameters were computed.

In order to eliminate user errors related to an excessively slow start of the selected type of movement, too rapid cessation of motion or incorrect initial movement, the learning datasets comprised 90 s recordings approx. in the middle of the 150 s time recording span.

According to the described recording procedure of the learning datasets (collected from 53 individuals), the following datasets were built:

1. For the procedure with signal parametrization: 900 42-element vectors for each individual. The calculations of these 900 vectors were made for each time window width $T \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 1.0, 2.0\}$, given in seconds, that were shifted with time steps of 0.1 s.
2. For the procedure based on the time domain: 9000 vectors for each individual, with six signal samples representing a "time capture" of the given head motion.

Finally, the test datasets on which the performance of individual classifiers were evaluated were correspondingly as follows:

1. For the *SP* procedure—set with 42-element vectors. Calculations for these testing vectors were made for time windows of widths $T \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8, 1.0, 2.0\}$ seconds, that were shifted with time steps of 0.1 s.
2. For the *TDR* procedure—set of vectors each with six signal samples.

The datasets are available for download at [29].

2.2. The Classifiers

For the purpose of classification of the head gestures on the basis of the recorded IMU signals, data classifiers generally recognized to be very efficient were applied [30]. In particular, various architectures of decision trees and decision forests were used since they can provide an insight into the data decision process, i.e., these classifiers can decompose the classification task into decision rules of the input data features. The data classifiers that were applied were the following: (1) a decision tree, (2) a decision tree with a minimum number of samples per leaf equal to 5, (3) a random forest consisting of 10 decision trees, and (4) a random forest consisting of 10 decision trees, each of which contained a minimum number of leaf samples equal to 5, (5) k-nearest neighbors method (k-NN) for $k \in \{1, 3, 5, 7, 9, 11, 13, 15, 19\}$, (6) Support Vector Machines (SVM) with a radial basis function kernel, and lastly (7) the SVM with a third degree polynomial as the kernel function.

2.2.1. Decision Trees and Random Forest

Decision tree is a supervised non-parametric classifier that allows building decision rules by multistage divisions of the dataset into disjoint classes [31].

Different algorithms can be used for building decision trees. In this work, we have employed the CART (Classification and Regression Trees) algorithm and used the so-called Gini coefficient as a measure of the diversity of classes in the tree nodes [25]. Classification trees can be combined into groups to form the so-called random forests. The final classification of the vector is performed by voting, i.e., a vector is assigned to the class for which it receives the largest number of votes from individual decision trees. Random forests are efficient classifiers for very large data sets [32,33].

2.2.2. The *k*-Nearest Neighbor classifier

The *k*-Nearest Neighbor Classifier (*k*-NN) is another non-parametric data classification technique. This is an instance-based learning algorithm that works on a simple scheme in which data sample x is assigned to a class for which there is a majority of data prototypes within the k nearest neighbors [34]. The k value should be selected in such a way as to retain the balance between overtraining and insufficient fit. We have evaluated the performance of this classifier for $k \in \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19\}$ neighbors. However, we present just the best classification results obtained for $k = 7$ for training procedure SP, i.e., with IMU signal parametrization and for $k = 19$ obtained for the time domain analysis. In all the tested *k*-NN classifiers, the Euclidean distance was used as the distance metric.

2.2.3. Support Vector Machines (SVM)

The SVM classifier is particularly suitable for classifying large dimensional datasets [35,36]. The basic optimization goal in this classifier is to maximize the margin, i.e., the between-class separation region. The margin is defined as the distance between the decision boundary (separation hyperspace) and the nearest learning samples, called the support vectors. The aim is to obtain decision boundaries with the widest possible margins. In our implementation of the SVM classifier for the multiclass problem, a scheme in which the one-vs-rest classification approach was adopted, we tested the performance of these classifiers for regularization parameter values $C \in \{0.1, 1.0, 5.0, 10.0\}$ for the two types of kernels, a RBF kernel and a 3rd degree polynomial. The best classification results were obtained for $C = 1.0$ for the parametrized signals and for $C = 10.0$ for time domain analysis. Thus, we present the classification results for the regularization parameter values for which the corresponding SVM classifiers performed best.

3. Classifier Training and Results

The classifiers were trained according to the procedures described in Section 2.1. The main aim was to compare the performance of different classifiers for parametrized signals and raw signal samples recorded from the inertial sensors.

The 10-fold cross validation method was used to evaluate the classifiers' performance in the head gesture classification task. The training sets (for both the *SP* and *TDR* training procedure) were therefore randomly divided into 10 different subsets of equal cardinality. Each subset was sequentially used as the validation set while the other nine subsets served as a training set for the classifiers. Each classifier was cross validated on the same subsets, i.e., the division of the training and validation sets was made first and then each classifier was trained on identical data subsets. Tables 1 and 2 report the results of the 10-fold cross-validation for the *SP* and *TDR* training procedures, respectively.

Table 1. Results of 10-fold cross-validation training—classification accuracy for the classifiers trained on the parametrized data consisting of 42 parameters (training procedure SP).

No.	Decision Tree	Cropped Decision Tree	Random Forest	Cropped Random Forest	*k*-NN for $k = 7$	SVM with RBF Kernel	SVM with 3rd Degree Polynomial
1.	97.63	97.84	98.57	98.52	95.40	97.82	98.14
2.	97.58	97.79	98.56	98.55	95.57	97.80	98.22
3.	97.50	97.74	98.64	98.58	95.51	97.86	98.22
4.	97.66	97.88	98.62	98.55	95.55	97.76	98.08
5.	97.55	97.87	98.71	98.60	95.52	97.68	98.17
6.	97.45	97.86	98.56	98.52	95.41	97.79	98.10
7.	97.49	97.81	98.57	98.57	95.50	97.77	98.03
8.	97.52	98.01	98.64	98.61	95.64	97.88	98.23
9.	97.62	97.86	98.65	98.55	95.49	97.84	98.17
10.	97.85	98.04	98.79	98.78	95.45	97.87	98.23
Mean ± Std	97.59 ± 0.11	97.87 ± 0.09	98.63 ± 0.07	98.58 ± 0.07	95.50 ± 0.07	97.81 ± 0.06	98.16 ± 0.07

Table 2. Results of 10-fold cross-validation training—classification accuracy for the classifiers trained on the IMU signal samples (training procedure *TDR*).

No.	Decision Tree	Cropped Decision Tree	Random Forest	Cropped Random Forest	*k*-NN for $k = 19$	SVM with RBF Kernel	SVM with 3rd Degree Polynomial
1.	97.45	97.59	98.17	98.05	98.19	94.27	92.58
2.	97.41	97.59	98.27	98.14	98.24	94.14	92.54
3.	97.48	97.57	98.26	98.13	98.21	94.31	92.59
4.	97.41	97.59	98.26	98.11	98.21	94.05	92.39
5.	97.41	97.52	98.18	98.06	98.20	94.25	92.53
6.	97.43	97.57	98.19	98.08	98.17	94.26	92.55
7.	97.45	97.60	98.27	98.10	98.19	94.28	92.53
8.	97.43	97.61	98.29	98.13	98.28	94.02	92.32
9.	97.42	97.63	98.23	98.11	98.22	94.14	92.41
10.	97.40	97.58	98.22	98.12	98.21	94.43	92.72
Mean ± Std	97.43 ± 0.02	97.59 ± 0.03	98.23 ± 0.04	98.11 ± 0.03	98.21 ± 0.03	94.22 ± 0.12	92.51 ± 0.11

For the classifiers trained on the parametrized signals (training procedure SP), the random forests and the SVM with a third degree polynomial kernel excelled and achieved accuracies over 98%. The poorest classification results were obtained for the *k*-NN classifier, but even this classifier achieved accuracies exceeding 95%.

For the classifiers trained directly on signal samples (training procedure *TDR*), all classifiers but the SVM achieved accuracies better than 97%. Although, the SVM classifier yielded results no worse than 92% correct head gesture recognitions.

We have evaluated the performance of the classifiers and compared whether the accuracies they yield are statistically different. Because we consider more than two classifiers and our data do not have a normal distribution, the Friedman test was used for this purpose [37]. For both the training procedures, i.e., the *SP* (with parametrization of IMU signals that generate 42 dimensional training vectors) and the *TDR* (raw IMU signal samples that generate six dimensional training vectors), we have rejected the null-hypothesis, i.e., we concluded that there are statistically significant differences ($p < 0.05$) between the classifiers. We also compared the chosen seven classifiers (Tables 1 and 2) by conducting the paired Wilcoxon signed rank test. For the *SP* training procedure for all classifiers we have also rejected the null-hypothesis. In the case of the *TDR* training procedure, there were no significant differences between the random forest and the *k*-NN for $k = 19$ (the null-hypothesis accepted). For other classifiers for the *TDR* procedure we have rejected the null-hypothesis, i.e., the classifiers' performances are statistically equal.

3.1. Results for Datasets Consisting of the Parametrised IMU Signals

For the parametrized signals (the *SP* training procedure), the classifier performances were evaluated for different time window widths T for which the statistical parameters of the signals were

calculated. The results obtained for the combined test scenarios $T1$ and $T2$ are shown in Figure 4 for different time window sizes. Also, Figure 5 compares the results of head gesture classifications obtained for two different testing datasets, i.e., the test data for trial participants whose data was used for training the classifiers and the test data from trial participant whose data was not used for training the classifiers. We can conclude that the results shown in Figure 5 show the user independent performance of the system.

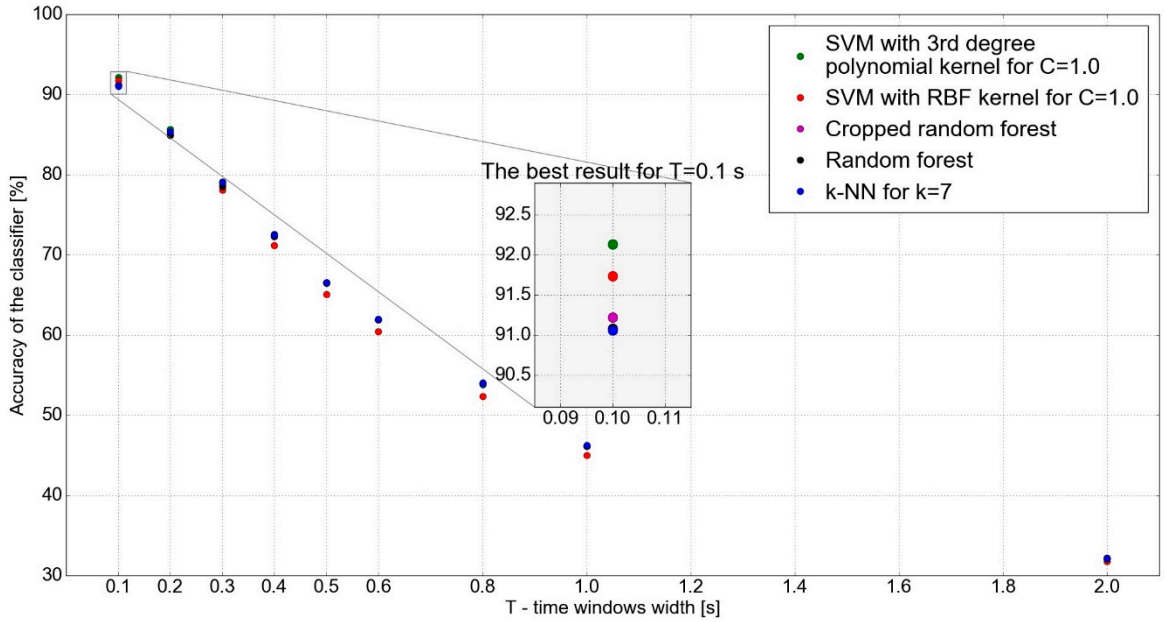


Figure 4. Accuracy of the classifiers for different widths of time window T for the test data.

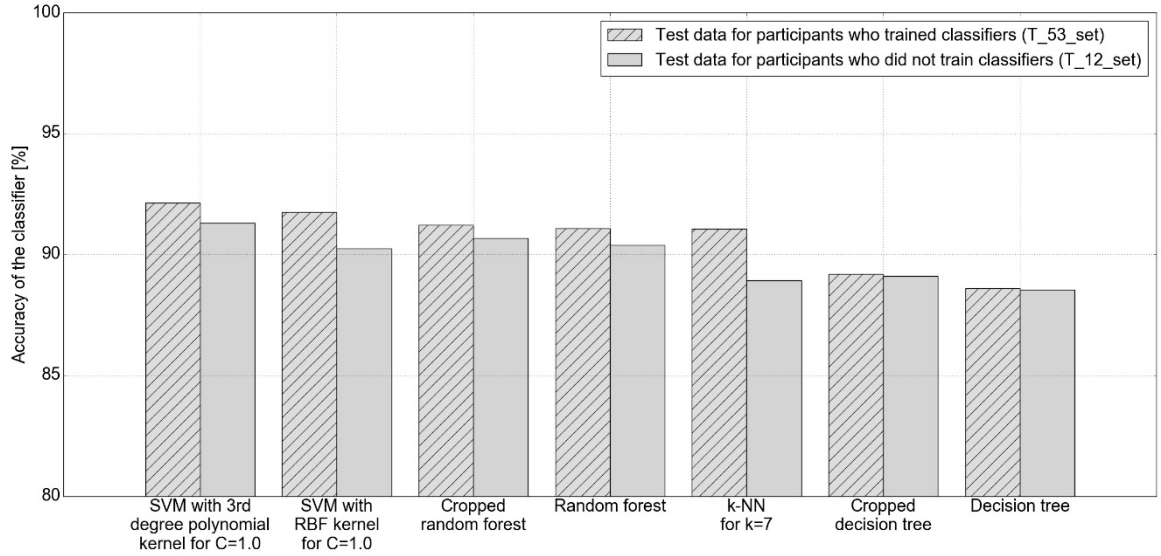


Figure 5. Comparison of the classifiers' accuracy obtained for the test data from participants who trained classifiers on the T_53_set and the test data from individuals who did not record the training data (T_12_set). Classifiers accuracy for the test data—causal system (vectors with 42 parameters).

Interestingly (as illustrated in Figure 4), the best results were obtained for the shortest time window, i.e., T = 0.1 s window containing 10 signal samples. An increase in the width of the time window negatively affected the classification accuracy. Note that for the test dataset, regardless of the width of the time window, the best classifier was the SVM with a 3rd degree polynomial kernel (accuracy better than 92%). The best classification accuracy for the k-NN classifier was obtained for

$k = 7$ and was slightly worse than the one achieved by the random forest classifier (both with an accuracy above 91%).

For the shortest time windows, the accuracies of individual classifiers tend to differ, whereas for larger time windows, the performance of all the classifiers converges to similar albeit poorer values. Figure 5 illustrates in more detail how the classifiers performed for the shortest time window ($T = 0.1$ s).

3.2. Results for Datasets Taken Directly from IMU Signal Samples

Subsequently, we evaluated the classifier performances for the testing datasets that were drawn directly from the samples of IMU signals. The achieved classifiers' accuracies for the test datasets are shown in Figure 6.

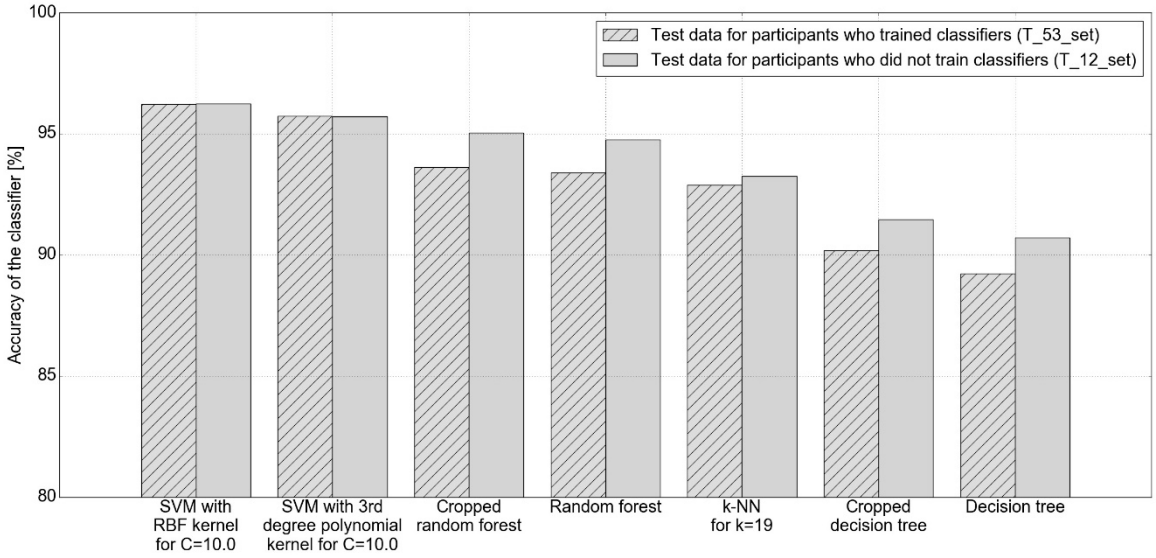


Figure 6. Comparison of the classifiers' accuracy obtained for the test data from participants who trained the classifiers (T_53_set) and the test data from individuals who did not record the training data (T_12_set). The test data collected directly from the accelerometer and gyroscope (vectors with 6 parameters).

Thus, we can note that for the classifiers trained on the IMU signal samples, the best accuracies exceeding 95% were achieved for the SVM classifiers. Also, an important observation is that their performance did not depend on the test data type (T_53_set or T_12_set).

3.3. Person Independent Recognition of Head Gestures

Here we report on the performance of the classifiers for the test data recorded for the 12 trial participants for whom the recorded IMU signals were not used for training the classifiers. In Figures 5 and 6, we compare the recognition rates of head gestures for the new 12 users and the classification results obtained for the test dataset recorded for 53 trial participants. Figure 5 contains the results for the shortest time window $T = 0.1$ s for which 42 statistical parameters were computed and used for training the classifiers, whereas Figure 6 shows the results for the classifiers trained on IMU signal samples, i.e., on six-element vectors.

For the person-independent test datasets, i.e., for the 12 participants who did not take part in training the classifiers, the best head movement recognition results for vectors with 42 parameters were obtained for the SVM and random forests. For all of these classifiers, about 90% accuracy or higher was achieved (Figure 5). Note also that for the person independent test datasets, the results obtained were inferior to those for the T_53_set test data (i.e., person dependent dataset) except for the SVM classifiers which performed equally on the two test datasets (for the *TDR* training procedure).

On the other hand, for the classifiers trained directly on the IMU signal samples, head gesture recognition accuracy exceeded 90% for all classifiers, except the decision tree classifier (see Figure 6). Interestingly, the accuracy for all classifiers except SVM was better for T_12_set, however, the SVM performed best for both T_53_set and T_12_set test datasets.

3.4. Head Gesture Recognition Rates for T1 and T2 Testing Scenarios

We also verified how the classifiers' accuracy depends on the type of testing scenario as defined in Section 2.1. For the causal system, the best results were obtained for test scenario *T*1 (see Figure 7). The accuracy of all classifiers exceeded 90%. The results obtained for test scenario *T*2 were distinctly inferior. For this testing scenario, the accuracy of the classifiers was below 90%. Again, the best classifiers were the SVM and the random forests.

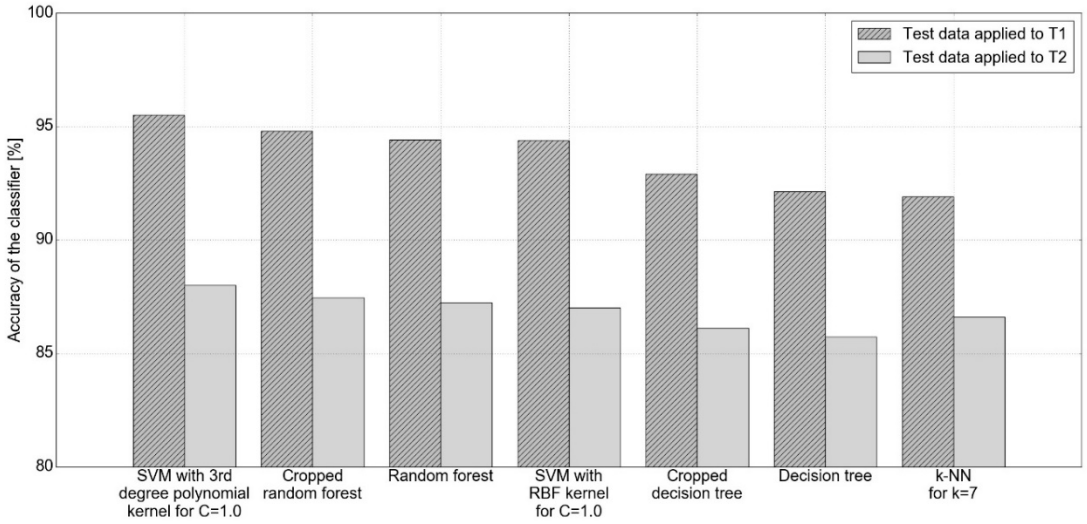


Figure 7. Comparison of the classifiers' accuracy obtained for test scenarios *T*1 and *T*2 for the test data recorded for individuals who did not record the training data (T_12_set) and the *SP* training procedure (classifiers trained on the parameters of the IMU signals).

For the signal samples recorded directly from the IMU (vectors with six parameters), almost all classifiers' accuracies were above 90%. As is evident in Figure 8, except for the SVM, the best results were obtained for test *T*2. Note that for the SVM classifier, the recognition accuracy for both tests was almost identical and exceeded 95%.

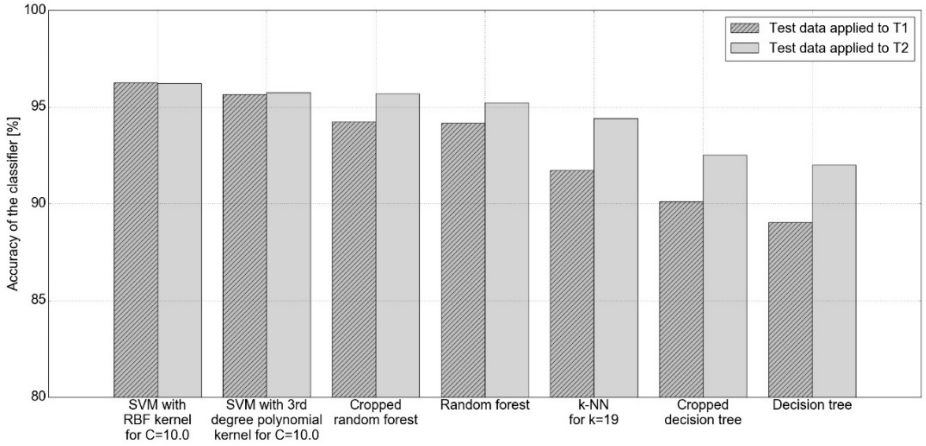


Figure 8. Comparison of the classifiers' accuracy obtained for test scenarios *T*1 and *T*2, i.e., for the test data recorder for individuals who did not record the training data (T_12_set). Classifiers accuracy for the test data collected directly from the accelerometer and gyroscope (vectors with six parameters).

We have also verified which head gestures were most often confused with one another for the classifiers which yielded the best results, i.e., for the SVM classifier with a 3rd degree polynomial and regularization parameter $C = 1.0$ for the *SP* training procedure and for the SVM with an RBF kernel and $C = 10.0$ for the *TDR* training procedure. The results are presented as confusion matrices in Tables 1 and 2. These tables show the results obtained for the 12 trial participants who did not take part in training the classifiers.

As shown in Table 3, the highest error rate was reported for pitch (5.5% of the gestures recognised as immobility). Further, considerably high error rates were noted for roll (4.6% of the gestures recognised as immobility). The best recognition rates were obtained for head immobility (92.6%). Interestingly, also for the immobility we have obtained the highest false positive recognitions.

Table 3. The confusion matrix showing inter-gesture recognition errors for the SVM classifier with a 3rd degree polynomial and $C = 1.0$ (the classifier trained on parametrised IMU signals).

	Recognised: Pitch	**Recognised: Roll**	**Recognised: Yaw**	**Recognised: Immobility**	
True: Pitch	92.0%	1.1%	1.4%	5.5%	100%
True: Roll	1.1%	91.7%	2.6%	4.6%	100%
True: Yaw	3.5%	1.8%	91.3%	3.4%	100%
True: Immobility	1.2%	2.8%	3.4%	92.6%	100%

The sensitivity, specificity and F1 scores of the head gesture recognition rates are presented in Table 4. Note that the F1 score is defined as a harmonic average of the sensitivity and positive predictive value. It is a good measure of overall classifier performance.

Table 4. Statistical measures showing detection results of head gestures for the SVM classifier with a 3rd degree polynomial and $C = 1.0$ (the classifier trained on parametrized IMU signals).

	Pitch	**Roll**	**Yaw**	**Immobility**
Sensitivity	92%	92%	91%	93%
Specificity	98%	98%	97%	96%
F1-score	92%	93%	92%	92%

Note that F1-score, sensitivity and specificity rates are similar for different gestures, however, the specificity rates assume the highest values.

For the SVM classifier trained on IMU signal samples (Table 5), the distribution of errors was different. For this case, the immobility was frequently and falsely recognized as yaw or as pitch (4.3% and 3.9% error rates respectively).

Table 5. The confusion matrix showing inter-gesture recognition errors for the SVM classifier with an RBF kernel and $C = 10.0$ (the classifier trained on IMU signal samples).

	Recognized: Pitch	**Recognized: Roll**	**Recognized: Yaw**	**Recognized: Immobility**	
True: Pitch	95.3%	0.3%	0.5%	3.9%	100%
True: Roll	0.3%	96.1%	1.8%	1.8%	100%
True: Yaw	0.6%	1.3%	93.8%	4.3%	100%
True: Immobility	0.5%	0.2%	0.5%	98.8%	100%

Note here that both the sensitivity and specificity rates in Table 6 (classifiers trained on raw IMU signals) outperform the corresponding rates in Table 4 (classifiers trained on parametrized IMU signals).

Table 6. Statistical measures showing detection results of head gestures for the SVM classifier with a 3rd degree polynomial and $C = 1.0$ (the classifier trained on raw IMU signals).

	Pitch	Roll	Yaw	Immobility
Sensitivity	95%	96%	94%	99%
Specificity	99%	99%	99%	97%
F1-score	97%	97%	95%	96%

Overall, however, the gesture recognition rates were better for the SVM classifier trained on IMU signal samples than the classifier trained on the parametrized IMU signals (compare the matrix diagonals in Tables 3 and 5, and also sensitivities, specificities and F1-scores in Tables 4 and 6).

The training times of the classifiers were also evaluated. The computations were performed on an Intel Core i5-7500, 16 GB RAM, Windows 10 64-bit PC. Scripts were written in Python in the Enthought Canopy environment using the Sklearn module. The obtained calculation times required for training the classifiers are presented in Table 7.

Table 7. Training times of the classifiers.

No.	Classifier	Parametrized IMU Signals	Raw IMU Signals
1.	Decision tree	12.6 s	18 s
2.	Decision tree with a minimum of 5 samples in the leaf	11.9 s	17 s
3.	Random forest	9.2 s	52 s
4.	Random forest with a minimum of 5 samples in the leaf	8.8 s	50 s
5.	k-NN for k ∈ {1, 3, 5, 7, 9, 11, 13, 15, 17, 19}	0.6 s	2 s
6.	SVM with an RBF kernel for C ∈ {0.1, 1, 5, 10}	5 min, 17 s	from 2 h, 27 min to 8 h, 42 min *
7.	SVM with a 3rd degree polynomial kernel for C ∈ {0.1, 1, 5, 10}	2 min, 12 s	from 2 h, 46 min to 9 h, 57 min *

* training times depend on the values of the regularization parameter *C*.

Depending on the types of datasets (42-element or 6-element vectors), a different size of the training datasets was used, hence the training times of the classifiers varied significantly. For the 42-element training vectors, the longest training time did not exceed 6 min. The training procedure was the fastest for the *k*-NN classifier. On the other hand, it took several minutes to train the SVM classifier.

For the 6-element training vectors (and larger training datasets), the shortest training times were obtained for the *k*-NN classifiers. Training of the decision trees (and random forests) took tens of seconds while the longest training times were required for the SVM classifiers (a few hours).

Table 8 shows the time necessary to recognize a head gesture from a single vector by the trained classifiers.

Table 8. Recognition times of a gesture for a single input data pattern.

No.	Classifier	Recognition of Parametrized Signals	Recognition of Raw IMU Signals
1.	Decision tree	~1 ms	~1 ms
2.	Decision tree with a minimum of 5 samples in the leaf	~1 ms	~1 ms
3.	Random forest	~10 ms	~6 ms
4.	Random forest with a minimum of 5 samples in the leaf	~6 ms	~7 ms
5.	*k*-NN for *k* ∈ {1, 3, 5, 7, 9, 11, 13, 15, 17, 19}	~5 ms	~2 ms
6.	SVM with an RBF kernel for C ∈ {0.1, 1, 5, 10}	~10 ms	~5–11 ms
7.	SVM with a 3rd degree polynomial kernel for C ∈ {0.1, 1, 5, 10}	~10 ms	~4–10 ms

The computation times required for processing the data by the trained classifiers (for a PC with Intel Core i5-7500 processor) varies from 1 ms for the decision tree classifier to 11 ms for the SVM classifier with an RBF kernel. It should be noted, that for the parametrized approach, we must first compute the statistical parameters; for the shortest time window consisting of 10 signal samples, we need to wait for 0.1 s before we feed these parameters to the classifiers.

4. Discussion

Our primary motivation for the study was to build a hands-free interface for the visually impaired that would facilitate controlling an electronic travel aid built in our project dedicated to help the visually disabled in independent mobility and travel [26]. The main role of the DUO MLX device and the image processing software is to recover a 3D structure of the environment from the captured stereovision images. The sequences of depth images are then sonified and acoustically presented to the visually impaired user. In [38], we have shown that such a non-visual presentation of the environment can help the blind to detect obstacles and find safe walking paths in the surroundings. One of the complications we encountered in the conducted trials is handling of the device by the blind. The use of an additional remote control to manipulate the interface proved to be inconvenient and occupied the other hand of the user who carried a white cane. An example scenario of how the interface might be used is the following. The navigation system generates a series of audio-haptic signals to indicate obstacles or inform about points of interest. Such a system features numerous settings like loudness, sensitivity range, sonification, and haptic activation schemes. The blind user, while standing still, by a series of head gestures, can move within the menu of the system and select the appropriate setting for the encountered environment or a navigation task at hand. System settings are confirmed by synthetized voice messages. It is important to note that a blind person carries a white cane which occupies one hand and the proposed solution does not need to engage the other hand to control the device. Also, after consulting the visually impaired, we plan another application of the interface. The device can play the role of a remote control that can aid the visually impaired in the activities of daily living, e.g., it can be used to control radio/TV volume without a need to seek for the remote of such devices.

Compared with the study reported in [24], we obtained a comparable F1-score of 92% for the parametrized IMU signals. It should be noted, however, that in [24] and in [17], individual gestures were not distinguished, but rather whole body activities, e.g., walking, running, cycling, hence the need for longer time data analysis windows: 5 s, 3 s and 1 s in [24] and 5 s and 12 s in [17]. In our case, we recognize delicate head movements, thereby the shorter the time window, the better the motion recognition. In [17,18], the IMU sensors were attached to the wrist, in [24] two to five sensors were used and they were mounted in different parts of the body, while in [19], IMU was mounted on the head like in our work. The advantage of our study is that 65 people participated in the experiments and our database is more diverse than e.g., in [19] where only five people took part in the trials. None of the works [17–19,24] compared the recognition results of body movements for data derived from the parametrized signals and time domain samples of IMU signals. We also tested the recognition rates from a large pool of different data classifiers, i.e., the decision trees and forests (and their cropped versions), k-NN, and SVMs (with an RBF kernel and a 3rd degree polynomial kernel). Finally, we show that better classification results can be obtained if the classifiers are trained on raw IMU signals rather than on the parametrized signals.

We are aware of the limitations of our study. The participants of the trials were mainly young people recruited from the students. Thus, an open question is on the usability of the system for elder people. Secondly, the trial participants remained seated during data collection. This limitation, for some system applications, should be released. Also, we should underline that our choice of IMU signals parameters, e.g., statistical parameters, was arbitrary and one might hypothesize that a different set of parameters might be composed that would even further improve the recognition performance of the head gestures.

Finally, the proposed system would need to be tested with target groups recruited from persons with physical or sensory disabilities.

5. Conclusions

In the presented study, we have compared two approaches to the classification of head gestures, i.e., training the classifiers on the parametrized signals, and training on direct signal samples recorded from the IMU.

The obtained results of head gestures recognition allow to formulate the following conclusions:

1. Head movements (*roll*, *yaw*, and *pitch*) can be efficiently recognized on the basis of signal recordings from an IMU that is positioned on the user's forehead.
2. The data classifiers trained on IMU signal samples outperformed the classifiers that were trained on a set of statistical parameters derived from the IMU signals. These performance differences were confirmed by running statistical tests. The Friedman test revealed that the classifiers yield statistically different results ($p < 0.05$). Also, the paired Wilcoxon signed-rank test has confirmed statistically valid differences for most of the classifiers (no significant differences between the random forest and the k-NN classifier for $k = 19$ was noted only). Our explanation about high head gesture recognition performance from raw IMU signals is that the six signal channels carry rich enough information for the classifiers to confidently recognize the head gestures. Also, we conclude that the inherent measurement noise (that occurs randomly) is averaged out during the training procedures of the classifiers.
3. The SVM classifier outperformed other classifiers in recognizing the head gestures, and if trained on the IMU signal samples, it achieved a recognition accuracy above 95%.
4. The recognition rates of the head gestures for the person-independent test dataset, i.e., the data taken from 12 persons whose recordings were not used in training the classifiers are comparable to the recognition rates obtained for the data recorded from the individuals who have also recorded the training data (see Figures 5 and 6). Interestingly, if the classifiers were trained on raw IMU signal samples, their recognition performance was generally better for data recorded from participants who did not train the classifiers.
5. The proposed method is suitable for on-line implementations. In Table 8, we show that the computation times required for detection of a gesture (on a PC with Intel Core i5 processor) by pre-trained classifiers do not exceed 11 ms.

We hypothesize that the presented head-gesture-controlled interface can find numerous applications. Firstly, it can offer an alternative communication channel for people with serious physical disabilities, e.g., individuals suffering from serious spinal injuries or stroke that result in tetraplegia. We made a new public benchmark dataset consisting of the training data (from 53 participants) and testing data (from 65 participants). The signal database with description is available for download at [29].

In further studies, on the basis of high head-gesture recognition performance, one could even further improve the recognition accuracy by accumulating classifier recognitions with each arriving IMU signal sample (i.e., with every 100 ms time interval) and achieve close to 100% recognition rates for the entire gesture. Our motivation is to further test the interface in a mobile travel aid system [38] and as a supporting interface in activities of daily living for the visually impaired, e.g., as a remote control system for home appliances. We will also consider the usefulness of such an interface as a potential rehabilitation aid for individuals with neck stiffness.

Author Contributions: A.B.-T.—data curation, formal analysis, methodology, software, visualization, writing—original draft. P.S.—conceptualization, methodology, project administration, supervision, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636 "Sound of Vision".

Acknowledgments: We thank all the participants for taking part in the trials reported in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. International Conferences on Human System Interactions, HIS. Available online: https://ieeexplore.ieee.org/xpl/conhome.jsp?punumber=1002118 (accessed on 29 May 2020).
2. Nielsen, J. *Usability Engineering*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1993.
3. Helal, A.; Mounir, M.; Abdulrazak, B. (Eds.) *The Engineering Handbook of Smart Technology for Aging, Disability, and Independence*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2008.
4. Strumillo, P.; Pajor, T. A vision-based head movement tracking system for human-computer interfacing. In Proceedings of the IEEE 2012 Joint Conference New Trends in Audio & Video and Signal Processing: Algorithms, Architectures, Arrangements and Applications (NTAV/SPA), Lodz, Poland, 27–29 September 2012.
5. Dumas, B.; Lalanne, D.; Oviatt, S. Multimodal Interfaces: A Survey of Principles, Models and Frameworks. In *Human Machine Interaction. Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2009; Volume 5440, pp. 3–26.
6. Poryzala, P.; Materka, A. Cluster analysis of CCA coefficients for robust detection of the asynchronous SSVEPs in brain–computer interfaces. *Biomed. Signal Process. Control* **2014**, *10*, 201–208. [CrossRef]
7. Kocejko, T.; Bujnowski, A.; Wtorek, J. Eye mouse for disabled. In Proceedings of the IEEE 2008 Conference on Human System Interaction (HIS), Krakow, Poland, 25–27 May 2008.
8. Krolak, A.; Strumillo, P. Eye-blink detection system for human–computer interaction. *Univers. Access Inf. Soc.* **2011**, *11*, 409–419. [CrossRef]
9. Ascari, R.; Silva, L.; Pereira, R. Personalized gestural interaction applied in a gesture interactive game-based approach for people with disabilities. In Proceedings of the International Conference on Intelligent User Interfaces, Cagliari, Italy, 17–20 March 2020; pp. 100–110.
10. Song, Y.; Luo, Y.; Lin, J. Detection of Movements of Head and Mouth to Provide Computer Access for Disabled. In Proceedings of the IEEE 2011 International Conference on Technologies and Applications of Artificial Intelligence, Chung-Li, Taiwan, 11–13 November 2011.
11. Jia, P.; Hu, H.H.; Lu, T.; Yuan, K. Head gesture recognition for hands-free control of an intelligent wheelchair. *Ind. Robot Int. J.* **2007**, *34*, 60–68. [CrossRef]
12. Solea, R.; Margarit, A.; Cernega, D.; Serbencu, A. Head movement control of powered wheelchair. In Proceedings of the 23rd International Conference on System Theory, Control and Computing, ICSTCC 2019, Sinaia, Romania, 9–11 October 2019; pp. 632–637.
13. Dobrea, M.; Dobrea, D.; Severin, I. A New Wearable System for Head Gesture Recognition Designed to Control an Intelligent Wheelchair. In Proceedings of the 2019 E-Health and Bioengineering Conference (EHB), Iasi, Romania, 21–23 November 2019; pp. 1–5. [CrossRef]
14. Ishizuka, A.; Yorozu, A.; Takahashi, M. Driving Control of a Powered Wheelchair Considering Uncertainty of Gaze Input in an Unknown Environment. *Appl. Sci.* **2018**, *8*, 267. [CrossRef]
15. Matsuzawa, K.; Ishii, C. Control of an electric wheelchair with a brain-computer interface headset. In Proceedings of the IEEE 2016 International Conference on Advanced Mechatronic Systems (ICAMechS), Melbourne, VIC, Australia, 30 November–3 December 2016.
16. Mitra, S.; Acharya, T. Gesture Recognition: A Survey. *IEEE Trans. Syst. Man Cybern. Part C* **2007**, *37*, 311–324. [CrossRef]
17. Yang, J.-Y.; Wang, J.-S.; Chen, Y.-P. Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers. *Pattern Recognit. Lett.* **2008**, *29*, 2213–2220. [CrossRef]
18. Maziewski, P.; Kupryjanow, A.; Kaszuba, K.; Czyzewski, A. Accelerometer signal pre-processing influence on human activity recognition. In Proceedings of the IEEE Signal Processing Algorithms, Architectures, Arrangements, and Applications SPA 2009, Poznan, Poland, 24–26 September 2009.
19. Wu, C.W.; Yang, H.Z.; Chen, Y.A.; Ensa, B.; Ren, Y.; Tseng, Y.C. Applying machine learning to head gesture recognition using wearables. In Proceedings of the 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST), Taichung, Taiwan, 8–10 November 2017; pp. 436–440. [CrossRef]

20. Lorenzi, P.; Rao, R.; Romano, G.; Kita, A.; Irrera, F. Mobile Devices for the Real-time detection of specific human motion disorders. *IEEE Sens. J.* **2016**, *16*, 8220–8227.
21. Baraka, A.; Shaban, H.; Abou El-Nasr, M.; Attallah, O. Wearable Accelerometer and sEMG-Based Upper Limb BSN for Tele-Rehabilitation. *Appl. Sci.* **2019**, *9*, 2795. [CrossRef]
22. He, C.; Kazanzides, P.; Sen, H.T.; Kim, S.; Liu, Y. An Inertial and Optical Sensor Fusion Approach for Six Degree-of-Freedom Pose Estimation. *Sensors* **2015**, *15*, 16448–16465. [CrossRef] [PubMed]
23. Montero Quispe, K.G.; Sousa Lima, W.; Macêdo Batista, D.; Souto, E. MBOSS: A Symbolic Representation of Human Activity Recognition Using Mobile Sensors. *Sensors* **2018**, *18*, 4354. [CrossRef] [PubMed]
24. Allik, A.; Pilt, K.; Karai, D.; Fridolin, I.; Leier, M.; Jervan, G. Optimization of Physical Activity Recognition for Real-Time Wearable Systems: Effect of Window Length, Sampling Frequency and Number of Features. *Appl. Sci.* **2019**, *9*, 4833. [CrossRef]
25. Van der Straaten, R.; Bruijnes, A.K.B.D.; Vanwanseele, B.; Jonkers, I.; De Baets, L.; Timmermans, A. Reliability and Agreement of 3D Trunk and Lower Extremity Movement Analysis by Means of Inertial Sensor Technology for Unipodal and Bipodal Tasks. *Sensors* **2019**, *19*, 141. [CrossRef] [PubMed]
26. Skulimowski, P.; Owczarek, M.; Radecki, A.; Bujacz, M.; Rzeszotarski, D.; Strumillo, P. Interactive Sonification of U-depth Images in a Navigation Aid for the Visually Impaired. *J. Multimodal User Interfaces* **2018**. [CrossRef]
27. Baranski, P.; Strumillo, P. Emphatic trials of a teleassistance system for the visually impaired. *J. Med. Imaging Health Inform.* **2015**, *5*, 1640–1651. [CrossRef]
28. The Producer's DUO MLX. Available online: https://duo3d.com/ (accessed on 29 May 2020).
29. Anna Borowska-Terka. Available online: http://eletel.p.lodz.pl/abterka (accessed on 29 May 2020).
30. Theodoridis, S.; Koutroumbas, K. *Pattern Recognition*, 4th ed.; Academic Press, Inc.: Orlando, FL, USA, 2008.
31. Koronacki, J.; Cwik, J. *Statistical Learning Systems*, 2nd ed.; Academic Publishing House EXIT: Warsaw, Poland, 2015. (In Polish)
32. Random Forests. Available online: https://www.stat.berkeley.edu/~{}breiman/RandomForests/cc_home.htm (accessed on 29 May 2020).
33. Zakariah, M. Classification of large datasets using Random Forest Algorithm in various applications: Survey. *Int. J. Eng. Innov. Technol.* **2014**, *4*, 189–198.
34. Aha, D.W.; Kibler, D.; Albert, M.K. Instance-based learning algorithms. *Mach. Learn.* **1991**, *6*, 37–66. [CrossRef]
35. Poulet, F.; Do, T.N. Mining Very Large Datasets with Support Vector Machine Algorithms. In *Enterprise Information Systems V*; Camp, O., Filipe, J.B.L., Hammoudi, S., Piattini, M., Eds.; Springer: Dordrecht, The Netherlands, 2004; pp. 177–184.
36. Schölkopf, B.; Smola, A.J. *Learning with Kernels*; MIT Press: Cambridge, MA, USA, 2002.
37. Janez, D. Statistical Comparisons of Classifiers over Multiple Data Sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.
38. Caraiman, S.; Morar, A.; Owczarek Burlacu, A.; Rzeszotarski, D.; Botezatu, N.; Herghelegiu, P.; Moldoveanu, F.; Strumillo, P.; Moldoveanu, A. Computer Vision for the Visually Impaired: The Sound of Vision System. In Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; pp. 1480–1489.