A classification and calibration procedure for gesture specific home-based therapy exercise in young people with cerebral palsy

Alexander Macintosh, Nicolas Vignais, Eric Desailly, Elaine A. Biddiss, Vincent Vigneron

A Classification and Calibration Procedure for Gesture Specific Home-Based Therapy Exercise in Young People With Cerebral Palsy

Alexander MacIntosh, Nicolas Vignais, Eric Desailly, Elaine Biddiss, and Vincent Vigneron

***Abstract*— Movement-based video games can provide engaging practice for repetitive therapeutic gestures towards improving manual ability in youth with cerebral palsy (CP). However, home-based gesture calibration and classification is needed to personalize therapy and ensure an optimal challenge point. Nineteen youth with CP controlled a video game during a 4-week home-based intervention using therapeutic hand gestures detected via electromyography and inertial sensors. The in-game calibration and classification procedure selects the most discriminating, person-specific features using random forest classification. Then, a support vector machine is trained with this feature subset for in-game interaction. The procedure uses features intended to be sensitive to signs of CP and leverages directional statistics to characterize muscle activity around the forearm. Home-based calibration showed good agreement with video verified ground truths (0.86 ± 0.11, 95%CI = 0.93-0.97). Across participants, classifier performance (F1-score) for the primary therapeutic gesture was 0.90 ± 0.05 (95%CI = 0.87-0.92) and, for the secondary gesture, 0.82 ± 0.09 (95%CI = 0.77-0.86). Features sensitive to signs of CP were significant contributors to classification and correlated to wrist extension improvement and increased practice time. This study contributes insights for classifying gestures in people with CP and demonstrates a new gesture controller to facilitate home-based therapy gaming.**

***Index Terms*— Cerebral palsy, exercise therapy, game, gestures, young adult, machine learning.**

Manuscript received June 12, 2020; revised November 6, 2020; accepted November 11, 2020.This work was supported in part by the Canadian Institutes of Health Research grant number: RN304779–379428. *(Corresponding author: Alexander MacIntosh.)*

Alexander MacIntosh is with the Department of Biomedical Engineering, University of Toronto, Toronto, ON M5S 1A1, Canada (e-mail: alexander.macintosh@mail.utoronto.ca).

Nicolas Vignais is with the CIAMS Department, Université Paris-Sud, 91400 Orsay, France, and also with the Department of Sport Sciences, Université Paris-Saclay, 91405 Orsay Cedex, France.

Eric Desailly is with the Fondation Ellen Poidatz, Recherche and Innovation Department, 77310 Saint Fargeau-Ponthierry, France.

Elaine Biddiss is with the Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, ON M4G 1R8, Canada, and also with the Department of Biomedical Engineering, University of Toronto, Toronto, ON M5S 1A1, Canada.

Vincent Vigneron is with the IBISC Department, EA4526, Univ-Evry, Université Paris-Saclay, 91405 Orsay Cedex, France.

Digital Object Identifier 10.1109/TNSRE.2020.3038370

I. Introduction

HOME-BASED rehabilitation exercises can augment traditional therapy and lead to improved performance of daily activities for people with disabilities [1]. Cerebral palsy (CP) is a disability due to injury or abnormality of the brain impacting 2.11 in 1000 live births [2]. Occurring near birth and persisting through adulthood, CP presents with positive and negative motor signs including spasticity, weakness, impaired selective control and sensory deficits [3]. To improve manual ability, frequent and intense practice of specific, therapeutic hand movements is recommended along with goal or activity-directed tasks. Advances in gesture recognition offers a way to practice these movements through engaging virtual environments (*e.g.* as a video game controller) and improve home-based training efficacy and engagement.

In typically developing populations, single channel amplitude-based features have most commonly been applied to gesture control. These features usually include mean absolute value, zero crossing, root mean square, variance and Wilson amplitude [4]. Classification of these types of features are often done with support vector machines, decision trees, k-means clustering and hidden Markov models and increasingly neural networks. With these methods, many gestures (>10 [5]) can be classified with high accuracy (>95%) using aggregated data and large training sets [6]. However, in a clinical population a personalized approach is necessary to address the individual's abilities, therapy goals and movement strategies. Accordingly, *home setup* of therapeutic movement practice through gesture recognition has remained challenging. Further, to facilitate *therapeutic practice*, myoelectric patterns are preferred here over optical inputs as myoelectric patterns can support practice even when someone is capable of only small and inconsistent gestures. People with CP may have noisy neurological commands when gesturing [7]. The muscle command to generate a targeted movement (*e.g.* hand opening) can be accompanied by: inconsistent neural drive (spasticity or weakness), atypical forearm flexor and extensor muscles synergies (undesirable co-contraction), and movement artifact (impaired selective control) [3], [6]. For these reasons, muscle activity has usually been conveyed as a line or character that moves in direct proportion with the change in activity or co-contraction [8]–[11] and researchers have been deterred from home-based gesture classification of therapeutic activities for the hand using commercial sensors for young people with CP. Using features specifically targeted to discriminating signs of CP may facilitate home-based gesture classification and calibration. This approach has the added potential benefit of providing *clinical insight*. In traditional home-based therapy activities, the precise quantity and quality of exercise repetition is unknown. However, with gesture-based activities discriminated by features of CP, clinicians may review the activity logs, determine the extent to which certain neuromotor signs persist and advise accordingly.

Finally, there are practical implications to consider when designing home-based therapy for young people. Laborious system configuration can challenge young people's attention, reducing motivation and engagement. To address this, we have previously defined design requirements by consulting with clinicians and persons with CP [12]. The system needs to be affordable, address a specific therapeutic movement goal, have simple hardware, quick semi-automatic gesture training, be embedded in an activity that keeps the individual's interest, and provide high quality feedback to both users and clinicians.

A. Aim

This study demonstrates an in-home calibration and classification procedure embedded into a rehabilitation video game for young people with CP. The procedure allows for:

a. Home setup: training data are collected and processed with minimal adult involvement within the first minutes of gameplay.

b. Therapeutic practice: Real-time gesture recognition identifies therapeutic movements for game feedback and control.

c. Clinical insight: Features used for classification are associated with neuromotor signs of CP.

The article is organized as follows. Section II A. details the software and hardware components of the system. Section II B. explains the home-based calibration and classification procedure. Section II C. contains information on 19 participants who completed the one-month intervention. Section II D. describes the analysis. Section III show the results of the analysis as they pertain to each aim of the study (a-c). In section IV the procedure, intervention and results are discussed, concluding with limitations to the current study.

II. Methods

A. Requirements for Use

The system requires hardware: laptop (typically used machine specifications: 2.67GHz CPU, Intel Core i5-560M, Intel HD Graphics, 4GB RAM), electromyography (EMG) and inertial sensor (Myo Armband, sampling at 200 Hz and 50 Hz respectively [13]) and software: adapted commercial video game (Dashy Square) and custom controller to interpret movements and command the game (MATLAB 2017b). Briefly, the objective of the game is to avoid all obstacles while the character progresses across the level. In the original commercial version, players tapped the screen at the correct moment to avoid obstacles. For this work, the tap command was replaced with the gesture controller. Level difficulty increased by requiring faster and more precise timing. Example gameplay can be seen at the link: here. Complete description of the commercial game adaptation are in an associated publication [12]. Before playing, participants were instructed of the game objective and coached by their therapist on how to perform the therapeutic gesture. To accommodate smaller arms, the original 8-channel device was cut to 4 or 6 channels and the software adjusts accordingly (see code on: GitHub). This reduction does not introduce any bias since the collected signals are statistically independent of each other, whether there are 4, 6 or 8 channels [14].

B. Home Data Processing Schema

Fig. 1 outlines how participant's data are collected and used in this home-based rehabilitation video game. A detailed description of each step is given.

1) Part 1 – Processed Before Game Play: Before gestures can be used to control the game, three steps are completed: calibration, feature selection, and building the online classifier. Methods for these steps are described next.

a) Calibration: Calibration is done at home, at the beginning of each session and appears to users as part of the game. To begin, the participant sits in front of the laptop and puts the sensor on the thickest part of the forearm. The participant launches the game and sits comfortably with the hand rested. They were allowed to choose to place the hand either on their lap or on the table next to the laptop. They were instructed to not let the sensor hit the table when possible. The game launches immediately into calibration. Calibration serves to 1. determine armband orientation, 2. initialize baseline flexor and extensor muscle activity, 3. collect training data for up to three gestures. To calibrate, participants follow animation prompts on-screen. The prompt tells them to keep a rested hand position for 3-seconds, then transition to a gesture and hold for 15-seconds. Doing so animates the character (video example link: here). Then the character stops moving, and they are promoted to rest again before doing the next gesture. This is repeated for each gesture." The actions take one minute in total. Following calibration, these data are pre-processed to: A) Determine armband orientation. The channel with the highest mean is designated as the primary extensor sensor. The flexor sensor is that furthest away from the extensor. B) Initialize baseline and maximum activity. The 25th percentile of the rest-phase data is set as the baseline. Similarly, the 80th percentile of the local maxima is used as initial maximum activity. These percentiles can be adjustable to the participant and were initially determined during preliminary testing [12]. This testing showed that outside these ranges, gestures were either erroneously considered attempted even during smallest arm movements or participants had difficulty repeatedly extending with sufficient intensity to elicit a command. Finally, data are prepare for feature selection. Here, raw 8-bit EMG calibration data are normalized to a 0-1 scale and windowed into 200ms bins without overlap. This bin size was selected through design

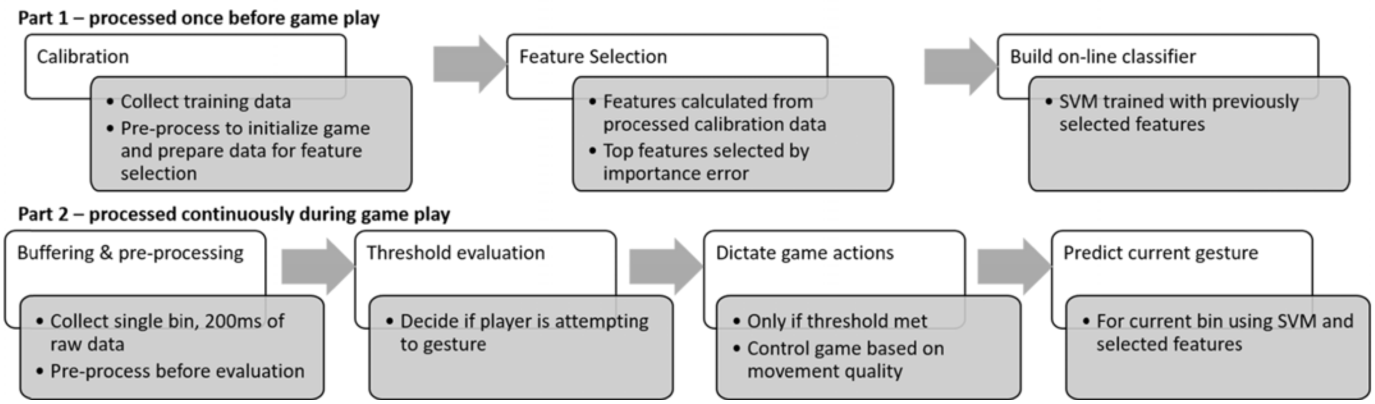

Fig. 1. Overview of calibration and classification procedure to use EMG and inertial data as controller and feedback in rehabilitation video game. Phases are divided into steps required before and during game play. Full details of each step are described in corresponding sections with the same name.

session testing conducted before the home-based trial and used to develop the system. In development, this was the largest window length consistently undetectable by participants while playing [12]. Window lengths between 50-300ms with overlap at 25, 50 and 75% were tested. No overlap value showed significant differences in classification performance and window lengths less than 150ms showed slightly lower performance [12]. Features described in the next step are calculated for each of these bins.

b) Feature selection: The most important features, derived from the EMG and inertial data collected with the Myo Armband, are selected via a 200 bootstrap-aggregated (bagged) random forest decision tree algorithm (MATLAB Treebagger [15]). Preliminary tests were made using 5-15 features and showed that after ten features, classification performance improvements were marginal. This variable is adjustable in the software. The measure of importance is out-of-bag permuted predicted delta error. In this method, feature importance [16] is evaluated as the difference in test set (out-of-bag) model error when the value of a feature is randomly permuted. Permuting the value of an influential feature indicates it is highly influential. This process is repeated at each tree for each feature [17]. Before selection, highly correlated (>0.8) redundant features were removed.

Therapists and participants were consulted to determine the targeted movements to practice in the game. Based on these targeted movements, features were selected. Preliminary testing showed that under certain signs of CP, *e.g.* high flexor tone, movement variability, extensor weakness, there can be small differences between gestures when using traditional EMG features [18]–[20]. We have therefore also included features expected to be sensitive to two signs of CP. Formulas for all features can be found in the Appendix and on GitHub. Features are put into three groups to facilitate clinical interpretation.

1. *CI group* – the Channel-Independent features, traditionally applied in typical developing populations or in prosthetic control. Features in this group include: Root mean square (RMS), Mean absolute value (MAV), Variance (VAR), Waveform length (WL), Zero crossing rate (ZC), Willison amplitude (WAMP), Slope sign changes (SSC) [18]–[20].
2. *RA group* - we considered features sensitive to uncommon Relative muscle Activities *i.e.* weakness, spasms, increased flexor tone, and impaired co-contraction [7]. The domain specific features in this group are: Co-contraction index (CCI), Scaled co-contraction index (SCCI), Mean absolute difference normalized (MADN), Scaled mean absolute value (SMAV) [21].
3. MV group - we considered features related to Movement Variability which may be associated with impaired selective motor control and compensatory movements [7]. To capture muscle activity precision around the forearm we employ directional statistics [22]. Directional statistics characterize the (Von Mises) distribution of muscle activity around the forearm as a unit circle. Features in this group include: Circular- mean, resultant, skew, kurtosis, standard deviation and variance [22]. To our knowledge, these features are new to EMG-based gesture classification and could offer a visual link to therapeutic movements via game control. Additionally, gross forearm movement was addressed via inertial features (squared sum of 3-axis acceleration/gyroscopic variability and magnitude).

c) Build on-line classifier: A support vector machine (SVM) was trained with the selected features [23]. The model was optimized with 5-fold cross-validation and hyperparameters (box constraint, kernel scale, kernel function, and polynomial order) were tuned via grid search [24]. The model is then used to predict gestures during gameplay. During preliminary design, alternative classifiers were tested (linear discriminant analysis, nearest neighbor, ensembles and naïve Bayes) details described in Section II – Analysis.

2) Part 2 – Processed During Game Play: After game values are initialized and the on-line classifier is prepared, participants can practice the targeted therapeutic gesture within the video game. EMG and inertial data from the Myo Armband are processed continuously as follows:

a) Buffering & pre-processing: Absolute values of the raw 8-bit EMG data from the Myo Armband are normalized to a 0-1 scale. For inertial data, the squared sum is calculated for the 3-axis accelerometer and gyroscopic data respectively. These are then buffered into a 200ms bin. This window length was determined during design session testing with youth with CP [7]. This was the maximum window length that could be stored before affecting sensitivity and control as perceived by participants.

Once data has buffered, it is pre-processed before game actions can occur. Pre-processing starts by taking the mean of each channel and scaling it to the current baseline and maximum range. Baseline and maximum values are determined by finding the local maxima/minima from the previous 30 seconds. The 80th percentile of the local maxima are used to avoid unrealistic transient maxima (*e.g.* sensor contact with table). Subsequently, the mean of the current bin of data is scaled to this local range. For instance, if the current baseline was 0.1, the current maximum was 0.9 and the mean of the current bin was 0.3, then the scaled value of the current bin would be (0.3-0.1) / (0.9-0.1) = 0.25.

Baseline and maximum values are updated every ten seconds. This time is programmatically adjustable but was determined based on user perceived sensitivity and control [7]. Iteratively updating the baseline and maximum values allow the controller to remain sensitive to the target therapeutic gesture while avoiding unrealistically high maximums caused by physical contact with the sensor (*e.g.* hitting the table accidently) or irrelevant activity (*e.g.* fist clenching or flexor spasms). Inertia data are not locally scaled as they do not directly dictate if game actions should occur. After data are scaled, they can be used to evaluate if the user is attempting a gesture and if game controls should be initiated.

b) Threshold evaluation: Once the mean of the current bin is pre-processed and scaled, it is compared to pre-defined thresholds. Under researcher supervision, thresholds are set weekly to adapt as the user progresses. To do this, all local maxima in the trial are identified. Design session testing conducted before the home-based trial and used to develop the system [12], revealed that when target gestures were attempted, flexor / extensor local maxima were consistently above the 10th percentile of all local maxima observed during the trial. Meanwhile, when the local flexor/extensor maxima were below the 10th percentile, the hand was at rest. The minimum local maxima value, after removing those below the 10th percentile was set as the threshold minimum. The 80th percentile was set as the threshold maximum to avoid unrealistic transient maxima caused by sensor contact. In this way, the activity threshold can be set unique to each gesture. If the current activities of the extensor and flexor channels are respectively above and below these thresholds, the user is considered to be attempting a gesture and data will be processed further to dictate game actions. If the thresholds are not met, processing stops until a new 200ms bin of data has buffered.

c) Dictate game actions: When the threshold is exceeded, game actions occur, *i.e.* players move, points are awarded, feedback is presented. Game actions are dictated by movement quality. Movement quality variables include: extensor-flexor co-contraction ratio, forearm angular acceleration and the predicted gesture. Based on these evaluations, commands are passed to the game and points are awarded.

d) Predict current gesture: To obtain the predicted gesture for the current 200ms bin of data, first the top 10 features are calculated (as established in Part 1 – processing before game play). EMG features are processed through a Bayesian recursive filter [25]. This reduces noise for on-line prediction while remaining sensitive to rapid changes as expected to control the game. Then the SVM is used to predict the current gesture on-line. This information can be used as feedback (*e.g.* higher points scored for the correct gesture) or as a control mechanism for the game (*e.g.* to execute a binary operation such as to jump over an obstacle).

Once game actions are complete and the current gesture is predicted, game play processing repeats with the newest buffered 200ms bin of data.

C. Participant Information

Nineteen (19) young people with CP completed a 4-week home-based intervention in France and Canada. Inclusion criteria were: mild-moderately impaired use of one hand (Manual abilities classification system (MACS) levels I-II [26]), able to follow simple instructions and no history of unmanaged epilepsy, no botulinum toxin treatment within 3 months or constraint-based movement therapy within 6 months. As the objective of this article is to detail the calibration and classification procedure, complete study methodology and clinical findings are presented in an associated publication [27]. A detailed description of the game design and an evaluation of the biofeedback provided can also be found in a complementary article [12]. Approval was obtained by Holland Bloorview's Research Ethics Board (#18-785) and the French Comité de Protection des Personnes (CPP, #2018-A00536-49). Caregivers gave written informed consent.

During the game, the participants sat at a table in front of the laptop with the elbow at 90 degrees and the palm facing down. They controlled the game using one of the following therapeutic gestures: wrist extension-open fingers, wrist extension-closed fingers, finger-thumb pinch, or supination (video of gestures can be seen at the link: here). Participants and therapists decided together which gesture to practice and set the practice schedule. Most (17/19) participants practiced wrist extension (open or closed) for the entire activity. In all cases, participants aimed to keep the wrist in an extended or neutral position while performing the gesture [27].

D. Analysis

Descriptive summaries of participant characteristics are provided first. Gestures were verified by visually labelling videos collected weekly by the researcher during a game play session. A visual signal emitted from the game software synchronized video, EMG and inertial data. Initiation and termination of each gesture were labeled manually (True Labels) as a ground truth. The number of True Label samples are provided for each participant in Table I.

1) First Aim, Home Setup: Two measures were used to show that training data could be collected, processed and used within the first minutes of gameplay. First, we use the

TABLE I
PARTICIPANT INFORMATION AND SAMPLE SIZE BY CLASS

ID	MACS	Control gestures	N samples Ext-O	Ext-C	Pinch
A	I	Pinch / Ext-O	868	1069	298
B	I	Ext-O	1105	1276	380
C†	I	Ext-O	872	1683	377
D	I	Ext-O	485	1272	375
E	II	Ext-O	716	1089	210
F	I	Pinch / Ext-O	1099	803	1425
G	I	Ext-O	735	1554	319
H	II	Ext-O / Ext-C	1950	721	345
I	I	Ext-O	663	1224	248
J†*	II	Ext-O / Ext-C	215	877	124
K†	I	Ext-O	442	1676	38
L†	II	Ext-O	793	1653	191
M †	I	Ext-O	952	2756	566
N†	II	Ext-O / Ext-C	194	937	17
O	I	Ext-O	614	1177	442
P†	II	Ext-O / Ext-C	2033	1339	382
Q	I	Ext-O	966	1988	578
R	II	Ext-O	1466	720	358
S†	I	Ext-O	2763	1693	398

Abbreviations: wrist extension-open fingers (Ext-O), wrist extension-closed fingers (Ext-C), finger-thumb pinch (pinch), manual abilities classification system (MACS). † indicates participants with secondary diagnoses as reported by therapists including learning disabilities, attention deficit hyperactivity disorder, autism spectrum disorder. *N* are the number of 200ms raw data bins. 20% of *N* samples were held out for testing. The pinch gesture for participants K and N had < 100 samples and were excluded from analyses. Some participants started with one gesture and moved to another if they found it too difficult. * Participant J attempted control with supination but was unable to reliably coordinate the movement so continued with Ext-O / Ext-C after the first week. These data were removed from the analysis.

agreement (raw and Gwet's chance-adjusted index agreement coefficient [28], [29]) between the assumed gesture, prompted during calibration, and the true gesture, verified by video. Second, we report processing time for both the calibration phase and a single loop in-game.

2) Second Aim, Therapeutic Practice: When evaluating model performance, the classifier's role (*i.e.* to determine the control input or to inform biofeedback) in the system must be considered. *When used to inform biofeedback*, balanced measures are favorable since misclassified observations have a smaller, less direct effect to the user's experience. To this end, we chose the F1-score to consider precision and recall as it is less influenced by imbalanced data. Mathew's Correlation Coefficient (MCC) was reported as the most informative single score to establish the quality of a classifier prediction [30]. However, *when the classifier prediction is used to control the game*, correctly identifying the targeted therapeutic gesture is paramount. A misclassified, false negative observation of the target gesture frustrates and reduces the user's confidence in the system. Whereas, in the situation of a false positive, game play continues and there is less risk for negative effect. Accordingly, three measures (F1-Score, MCC, Sensitivity) are prioritized. Performance across participants for the whole dataset is presented for randomly partitioned off-line classification using the SVM procedure described above. To validate the use of SVM, multiple classifiers on a subsample of five participants were evaluated. Five participants were chosen to represent the distribution in functional ability and performance. Two were at MACS level II and three were younger than 12 years old. Tested classifiers include decision trees, linear discriminant analysis, nearest neighbor, ensembles and naïve Bayes [31]. The top two performing classifiers based on this subset are presented here. These are the SVM and the random subspace ensemble classification of k-nearest neighbor learners (ENS) (MATLAB fitensemble) [23]. Full model specification are available on GitHub. Reported also are weekly changes in the true positive and true negative rate (AUC) during home-based classification. AUC was computed across classes and then averaged across all participants. Here only calibration-game data up to the current week were used to train the model and tested against actual game-play data. To further evaluate the approach, performance of the SVM models with personalized features was compared to a model derived from all participants data combined (e.g. an 'All-Users' model).

3) Third Aim, Clinical Insight: First, we assess the value of feature groups expected to be sensitive to signs of CP (*uncommon relative muscle activities, RA-group,* and *high movement variability, MV-group)* as compared to traditional *channel-independent* features (*CI-group*). Each group's relative importance was calculated as the sum of the reciprocal rank of each feature within a feature group [32]. For instance, if features only within the *channel-independent* group were used for classification, that group's relative importance would be equal to one, and the remaining two groups would be zero). A one-way analysis of variance (ANOVA) was used to evaluate differences in relative feature group importance.

Data were approximately normally distributed as verified through visual inspection of standardized residuals and Shapiro-Wilks test (W = 0.97, p =0.25). Homogeneity of variance was verified by Levene's test (F = 1.22, p = 0.303). Post-hoc testing with Bonferroni adjustments for multiple comparisons were used to identify differences between groups [33]. α risk set to 0.05 was considered significant for all tests. The most important individual features and the variability of feature use across participants is also reported. As an exploratory objective, we evaluated linear correlation coefficients between the feature group importance and two clinical measures: a) wrist extension amplitude, and b) grip strength as related to the amount of practice and number of repetitions [12], [27].

III. Results

Table I shows participant information and the available data set size. Detailed demographics can be found in the complementary clinical publication [27]. Briefly, there were 19 participants (ten females, average age: 11.7 ± 2.5 years). Twelve participants had mildly impaired hand function and the remaining were categorized as moderately impaired (MACS level I and II). Practice dose over the one-month intervention averaged 4 ± 1 days/week (8 - 24 days total), 17 ± 9 minutes/day (37 - 333 minutes total), and 163 ± 59 gesture repetitions/day (997 - 5698 total) [20].

A. Aim 1 Home Setup

All but two participants were able to immediately follow the calibration prompts presented in-game. They were able to keep a rested state until prompted to make the required gesture and then relax at the end of the animation. Two participants (K, N) had particular difficulty following instructions. They would begin the gesture early and make excessive arm movements instead of only the required gesture. Agreement was high between the assumed gestures, collected during the calibration game, and the video verified true gesture. Most participants, 16/19 had >80% agreement. Agreement was also >80% for each class (Table II). Full setup time (including system start-up, calibration game, and entry to the main game) took 115.4 ± 144.2, 95%CI = 61.3 - 169.5 seconds. Of this, the time from the end of calibration to the start of gameplay averaged 9.1 ± 2.2, 95%CI = 8.2 - 9.9 seconds. During gameplay, on-line processing for each loop (200ms window of data) took 0.021 ± 0.008, 95%CI = 0.018 - 0.025 seconds when control thresholds were exceeded. In this time, features were calculated and classified to command the game. When thresholds were not exceeded, processing time was shorter (0.009 ± 0.003, 95%CI = 0.007 - 0.010 seconds). Loop times were rarely perceived by participants.

B. Aim 2 Therapeutic Practice

Overall participant specific classification accuracy in the primary target gesture, extension-open fingers averaged: 0.90 ± 0.05, 95%CI = 0.87 - 0.92 and 0.82 ± 0.09, 95%CI = 0.77 - 0.86 in the extension-closed fingers gesture and 0.80 ± 0.18, 95%CI = 0.72 - 0.89 for pinch gesture. Participant J used supination for only one week and therefore this data removed from the analysis. Table III shows F1, MCC and sensitivity scores by class and model. Overall accuracy was 4 ± 2 % higher in the Personalized model compared to the All-Users model. Table IV shows F1, MCC and Sensitivity scores by class and model (Personalized versus All-Users).

TABLE II
Reliability of In-Home Calibration During Gameplay

	Agreement			Gwet's AC1		
	Mean	**SD**	**95%CI**	**Mean**	**SD**	**95%CI**
Participant	0.86	0.11	0.93 - 0.97	0.83	0.14	0.93 - 0.97
Ext-C	0.90	0.12	0.88 - 0.92	0.88	0.15	0.86 - 0.91
Ext-O	0.87	0.14	0.85 - 0.89	0.83	0.20	0.81 - 0.86
Pinch	0.87	0.12	0.85 - 0.89	0.84	0.17	0.82 - 0.87

Home-based in-game calibration procedure agreement with video-verified ground truth. Mean and standard deviation (SD) with 95% confidence intervals (CI) across the 19 participants and each class: wrist extension-closed fingers (Ext-C), wrist extension-open fingers (Ext-O), finger-thumb pinch in neutral wrist posture (Pinch). Agreement shows raw value, and Gwet's AC1 shows chance-adjusted agreement index.

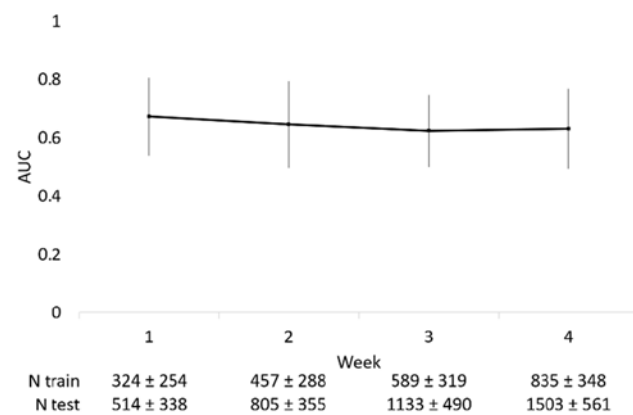

Fig. 2. Target class (wrist extension-open fingers (Ext-O)) AUC (area under curve of false positive rate * true positive rate) for each training week averaged across participants (N = 19). Sample sizes listed below corresponding weeks are cumulative mean ± standard deviation of training (N train) and testing (N test) set observations across participants.

Fig. 2 shows weekly home-based classification performance averaged across participants. Home-based classification was completed with training sets obtained during the 1-minute calibration. There was a disparity in performance across participants yielding high standard deviation. Specifically, low performance for three participants (mean AUC for participant: D = 0.43, N = 0.48, Q = 0.48). All remaining participants averaged an AUC of 0.67 ± 0.9.

C. Aim 3 Clinical Insight

There was an overall difference between feature group relative importance based on out-of-bag permuted predicted delta error (Fig.3, ($F_{2,54}$ = 8.08, $p < 0.001$). Post-hoc testing

TABLE III
GROUP-LEVEL CLASSIFICATION PERFORMANCE

	F1			MCC			Sensitivity		
	Mean	**SD**	**95%CI**	**Mean**	**SD**	**95%CI**	**Mean**	**SD**	**95%CI**
SVM									
Ext-C	0.76	0.12	0.71 – 0.82	0.66	0.13	0.59 – 0.72	0.75	0.15	0.68 – 0.83
Ext-O	0.87	0.06	0.84 – 0.90	0.72	0.11	0.66 – 0.77	0.88	0.07	0.84 – 0.91
Pinch	0.78	0.14	0.71 – 0.85	0.76	0.14	0.69 – 0.82	0.77	0.19	0.68 – 0.86
ENS									
Ext-C	0.82	0.09	0.77 – 0.86	0.73	0.10	0.68 – 0.78	0.80	0.14	0.73 – 0.87
Ext-O	0.90	0.05	0.87 – 0.92	0.77	0.10	0.72 – 0.81	0.91	0.07	0.87 – 0.94
Pinch	0.80	0.18	0.72 – 0.89	0.80	0.15	0.73 – 0.87	0.75	0.23	0.64 – 0.86

Classifier performance mean and standard deviation (SD) with 95% confidence intervals (CI) across participants (N=19) in F1-Score (F1), Mathew's Correlation Coefficient (MCC), and Sensitivity. Two model results shown: support vector machine (SVM) and ensemble classification by random subspace of k-nearest neighbors (ENS). These were the two highest performing of all tested classifiers. The classes were: wrist extension-closed fingers (Ext-C), wrist extension-open fingers (Ext-O), finger-thumb pinch in neutral wrist posture (Pinch).

TABLE IV
GROUP-LEVEL CLASSIFICATION PERFORMANCE PERSONALIZED VERSUS ALL-USERS MODELS

	F1			MCC			Sensitivity		
	Mean	**SD**	**95%CI**	**Mean**	**SD**	**95%CI**	**Mean**	**SD**	**95%CI**
Personalized									
Ext-C	0.76	0.12	0.71 – 0.82	0.66	0.13	0.59 – 0.72	0.75	0.15	0.68 – 0.83
Ext-O	0.87	0.06	0.84 – 0.90	0.72	0.11	0.66 – 0.77	0.88	0.07	0.84 – 0.91
Pinch	0.78	0.14	0.71 – 0.85	0.76	0.14	0.69 – 0.82	0.77	0.19	0.68 – 0.86
All-Users									
Ext-C	0.73	0.03	0.67 – 0.80	0.61	0.03	0.54 – 0.67	0.71	0.04	0.62 – 0.80
Ext-O	0.84	0.02	0.80 – 0.88	0.67	0.03	0.60 – 0.73	0.86	0.02	0.82 – 0.91
Pinch	0.74	0.04	0.66 – 0.82	0.71	0.04	0.62 – 0.80	0.70	0.04	0.61 – 0.80

Classifier performance mean and standard deviation (SD) with 95% confidence intervals (CI) across participants in F1-Score (F1), Mathew's Correlation Coefficient (MCC), and Sensitivity. Two model results shown based on features selected from participant-specific data (Personalized) and from all participants data combined (All-Users) models using support vector machine classification. The classes were: wrist extension-closed fingers (Ext-C), wrist extension-open fingers (Ext-O), finger-thumb pinch in neutral wrist posture (Pinch).

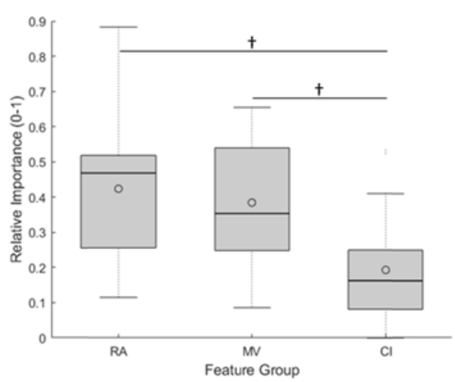

Fig. 3. Relative feature group importance showing *uncommon relative muscle activities* (RA) group and *high movement variability* (MV) group with greater importance than traditional, *channel-independent* (CI) feature group. † Indicates feature groups with significantly greater importance compared to the CI feature group ($p < 0.01$).

showed a lower mean importance in the channel-independent (CI) feature group (0.19 ± 0.16) compared to both the uncommon relative muscle activity (RA) feature group (0.42 ± 0.21, $p = 0.001$) and the high movement variability (MV) feature group (0.38 ± 0.19, $p = 0.009$). Of all the possible features ($N = 20$), the ten most common features were selected 80% (SD = 14, 95%CI: 38-94%) of the time. The ten most commonly used features in order of frequency were: SMAV_Ext, MADN_Ext, SMAV_Flx, MAV_Ext, AccMag, MADN_Flx, CCI, circMean and circSkew. The three most important individual features were the top features in 37% of occurrences across participants.

Table V shows correlations between the relative importance of each feature group and two clinical measures wrist extension amplitude and grip strength with respect to the average number of repetitions/session and the total practice time. In the accompanying clinical paper [27], outcomes for all measures of manual capacity are detailed.

IV. DISCUSSION

This study outlines an in-home calibration and classification procedure used for manual therapy activities of young people with CP. The procedure was personalized, tested in the homes of 19 families for one month and integrated into an adapted commercial video game. We show that participants followed the game-led calibration procedure with high agreement to ground truths. The classification methods employed facilitated practice of targeted therapeutic gestures. Finally, feature groups (RA and MV) expected to be sensitive to signs of CP were significant contributors and correlated with changes in practice and increased wrist extension capacity. This latter result strengthens the fact that not only traditional but also specific features should be considered when classifying movements from people with motor disabilities.

TABLE V
CORRELATION BETWEEN FEATURE GROUP IMPORTANCE, PRACTICE AND MANUAL CAPACITY

	RA			MV			CI		
	r	p	95%CI	r	p	95%CI	r	p	95%CI
Practice									
Reps/session*	0.39	0.10	-0.07 - 0.72	-0.43	0.06	-0.74 - 0.03	-0.01	0.95	-0.47 - 0.44
Practice time (min)	0.27	0.27	-0.21 - 0.64	-0.19	0.44	-0.59 - 0.29	-0.14	0.57	-0.56 - 0.34
Outcome									
Ext-O (deg)**	0.49	0.04	0.02 - 0.79	-0.22	0.39	-0.64 - 0.29	-0.40	0.11	-0.74 - 0.10
Grip strength (%)	0.12	0.64	-0.38 - 0.57	-0.36	0.16	-0.72 - 0.15	0.26	0.31	-0.25 - 0.66

Correlation coefficients (r), significance (p) and 95% confidence intervals (95%CI) between each feature group (*uncommon relative muscle activities* (RA), *high movement variability* (MV), and *channel-independent* (CI)), practice and outcomes (pre-post change measures of manual capacity: amplitude (joint angle, degrees) of active wrist extension- fingers open (Ext-O) and grip strength (non-dominant as percent of dominant)). * Indicates increased repetitions/session associated with decreased use of MV features and an increased use of RA features. ** Indicates greater improvements in wrist extension associated with increased use of RA features and decreased use of CI features.

A. Practical Implementation

Practical implementation is a key requirement of this procedure. While designed to be a fun, seamless experience blending game and therapy, the individual's attention on their movement is still critical. Our participants were 8-18 years old, some with learning disabilities or attention disorders, some with little general interest in video games. These factors influence how consistently one repeats their target gesture, especially when the gesture is difficult for them (*i.e.* a therapy goal). Accordingly, some participants (3/19) had difficulty following the calibration instructions and correspondingly low agreement scores and classification performance. Methods presented here do help implement home-based calibration and classification, but future work should ensure users can keep attention (at least 15-minutes) and display some consideration to how they perform a gesture (*i.e.* are they aware of the position/orientation of their hand?).

In addition to attention, out-of-laboratory myoelectric-based gesture recognition performance is known to be limited by interference including: electrode shift, muscle fatigue, unwanted motion and force variation [34]–[36]. To compensate for this, Ding *et al.* (2019) circumvent the burden of frequent retraining by using an adaptive incremental hybrid classifier [37]. The proposed method retrains target gestures in a semi-automated process by separating classes through resting and active states [37]. While manual input is still required, adaptive classifiers like this and others ([38]–[40]) may be a promising approach to address real-world EMG variability.

Compared to similar works with respect to sensor configuration and target gestures, classification performance was slightly lower. This can be expected given the unique population and that other studies concentrate on classifying more gestures with healthy adults, usually in the lab. For instance, k-nearest neighbor and dynamic time warping algorithms have been used with the Myo to achieve 86% accuracy across 5 [41]. However, this was completed with healthy participants in controlled situations. Similarly, 9 gesture classification in the lab using LDA found 9.86 $\pm$ 8.05% overall error [42]. To our knowledge, this is the first home-based gesture classification procedure for therapeutic activities for the hand using commercial sensors for young people with CP.

As these and other classification methods develop, new models can be implemented to improve the performance of this procedure. With this in mind, we present performance of an alternative classifier, ENS. Further, the software allows alternative classifiers to be called in place of SVM.

B. Feature Groups

Task specific muscle synergies in people with CP are influenced by neural and biomechanical factors including flexor spasms, weakness, and pathological reciprocal inhibition leading to unbalanced co-contraction. RA features characterize relative activity around the forearm and are person-dependent. The use of these features indicates that participants create differential flexor/extensor synergies between gestures. This is meaningful, as high flexor tone and unbalanced co-contraction often limits finger extension capacity, particularly when the wrist is above neutral [43]–[45]. Distinguishing this small difference (between open and closed finger extension) is an important quality to RA features and is reflected in their correlation to improvements in wrist extension.

Similarly, MV features characterize movement isolation through inertial measures and directional statistics of the forearm muscle activity [22]. In addition to significantly contributing to classification, these features were used as biofeedback in the game to help participants improve their ability to isolate movements at the wrist (a secondary clinical objective identified by occupational therapists). In this study, the average number of repetitions per session was correlated to a decreased use of MV features. This suggests that with increased exposure to the system, movement variability between gestures would decrease.

The frequent occurrence of three features (SMAV-Ext, MADN-Ext and SMAV-Flx) could be expected. First, given the target gestures the total activity of the extensor sensors (as represented by the feature SMAV-Ext) and the difference in activity between sensors on the extensor muscles and those adjacent (represented by MADN-Ext) would be defining characteristics of those actions. Second, to repeatedly produce these gestures, flexor muscle activity (SMAV-Flx) must to be consistent during each gesture. It is interesting to note that while among the most common, co-contraction indices were used in 6% of occurrences. Originally, it was expected that co-contraction would be more frequently used, since balancing extensor-flexor activity is key to opening and closing the hand. This lower proportion may be because co-contraction is a derivative of most common features SMAV-Ext and SMAV-Flx. Further, while ten features were used 80% of the time, It is worth noting that older (>12 years), less affected (MACS level I) participants used the top features on average 84% of the time, while younger, more affected participants used these top features only 62% of the time. This may indicate the greater difficulty with which these participants had in consistently producing target gestures.

We expect the method described in this study will be extensible to an increased number and complexity of gestures for two reasons. First, alternative, improved, classifiers can replace the SVM algorithm currently employed. Second, the features used here address the unique biomechanics of the hand. The feature types (particularly, circular features and relative activity features) are well suited to identify small differences in muscle activity across gestures. While the software can be used in its current form with modification, real-world deployment would be greatly improved by migrating to a more available language (e.g. Python) and improving the efficiency of the scripts. This work is currently being planned.

Clinically relevant feature groups were a key strategy we used to account for unique neuromuscular profiles. It should be noted that these features do not necessarily indicate severity of the disability nor does use of only traditional features indicate absence of any symptoms. Their benefit in this context is that these feature values fluctuate more between target and secondary gestures allowing for classification and feedback with respect to biomechanically relevant variables. Alternative strategies may also be effective. Kieliba *et al.*, (2018) used Factor Analysis to extract muscle synergies as features for classification [46]. These synergies can provide physiological plausible explanations of tasks and may be stable across participants and conditions [47]–[49]. However, these methods have yet to be tested on children with neuromuscular disabilities. It should be noted that feature selection improves performance while removing irrelevant or low variance characteristics, reducing the complexity of the model. A strength of the approach presented here is that feature selection was highly adaptable. Selected features were person-specific and updated weekly as the participant progressed. This allows for features of different channels to be included if they play a more prominent role in the target gesture and irrelevant features to be removed as gestures become more consistent.

C. Limitations

Weekly visits were used to monitor and adjust the activity thresholds and ensure calibrations were performed correctly. Thresholds were automatically adjusted weekly and manually verified by the researcher. For three participants who had greater difficulty following instructions during the practice level, thresholds were manually adjusted. In future work, calibration would need to be monitored by therapy staff or family members to ensure the game was facilitating practice of relevant movements. We recognize that alternative thresholds may be appropriate for different situations and as such have included this as a configurable item in the software.

The subsample of participants used to evaluate classifier performance, while selected to be representative of the variability in features used and gesture performance consistency, could have underestimated performance of some classifiers, not used within the game. Classifier performance may also be affected by window length and overlap. We acknowledge that alternative window lengths and overlap may be appropriate for different situations and as such have included this as a configurable item in the software. Additionally, future study is needed to determine the degree of sensitivity and accuracy in classification performance yield a noticeable improvement in user experience. Choosing the top ten features was selected as a reasonable compromise between on-line processing time and classifier performance. Minimal improvements were seen after ten features. However, this could be improved by using feature selection methods that choose the number of top features by maximizing prediction accuracy on a participant-specific basis, for instance Neighborhood Component Analysis (NCA) feature selection [50]. Further, deep learning methods such as convolutional neural networks may be promising towards providing more accurate predictions. These methods were not tested here since one goal was to provide clinical insight through features sensitive to neuromuscular signs of CP.

Ground truths were established manually via synchronized video evaluation. Distinguishing on video if the participant was extending with fingers open or closed was not always clear. Particularly when the person had minimal capacity to open the hand, how to label a frame can be somewhat subjective. Multiple raters would help improve confidence here. Further, closed versus open finger extension may be better evaluated on a continuum since gesturing with a partially opened hand still has merit towards therapeutic practice and is a signal of the user's intent. Further, differences between how the calibration game and the main activity are played could yield different muscle activities and feature distributions. Minimizing and accounting for the shift between training and real-world data requires continued investigation.

Finally, the target gesture for most participants was wrist extension. As such, the main objective in the game was to distinguish between extension with open and closed fingers. Some participants identified pinching as a goal. As such, pinching was included in the weekly calibration but rarely used in-game. This resulted in substantially lower training and testing data sets and lower classifier performance for this class. While the calibration procedure supports three classes. The in-game performance was based on successfully identifying open-finger extension. Future work should extend the on-line controller to three classes, such that each gesture controls a separate action in the game. The system and game are currently capable of this, but it has yet to be comprehensively tested at-home.

Appendix
Feature Formulae

A. Channel – Independent Group

Root mean square (RMS)

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x_n^2} \tag{1}$$

where x_n represents the EMG signal in a segment and N denotes the length of the EMG signal.

Mean absolute value (MAV)

$$MAV = \frac{1}{N}\sum_{n=1}^{N} |x_n| \tag{2}$$

Variance (VAR)

$$VAR = \frac{1}{N-1}\sum_{n=1}^{N} x_n^2 \tag{3}$$

Waveform length (WL)

$$WL = \sum_{n=1}^{N-1} |x_{n+1} - x_n| \tag{4}$$

Zero Crossing Rate (ZC)

$$ZC = \sum_{n=1}^{N-1} \left[\operatorname{sgn}(x_n \times x_{n+1})\, |x_n - x_{n+1}| > threshold\right];$$
$$\operatorname{sgn}(x) = \begin{cases} 1, & if\ x \geq threshold \\ 0, & otherwise \end{cases} \tag{5}$$

The threshold condition is used to avoid background noise of the signal, *sgn,* and evaluate ZC after given minimal activity (mean of, n in the current study).

Willison amplitude (WAMP)

$$WAMP = \sum_{n=1}^{N-1} f(|x_n + x_{n+1}|);$$
$$f(x) = \begin{cases} 1, & if\ x \geq threshold \\ 0, & otherwise \end{cases} \tag{6}$$

Slope sign changes (SSC)

$$SSC = \sum_{n=1}^{N-1} \left[f\left[(x_n - x_{n-1}) \times (x_n - x_{n+1})\right]\right];$$
$$f(x) = \begin{cases} 1, & if\ x \geq threshold \\ 0, & otherwise \end{cases} \tag{7}$$

B. Relative Activity Group

Mean Absolute Difference of the Normalized (MADN)

$$MADN_i = \frac{\sum_{n=1}^{wl} |x_i[n] - x_{i+1}[n]|}{\sum_{n=1}^{wl} x_i[n]} \tag{8}$$

where $x_i[n]$ is the n^{th} data point from channel i after the data in the window is normalized (mean value subtracted from each raw data point, and then the resulting values are divided by their standard deviation). wl is the window length, or the number of raw data points in one window.

Scaled Mean Absolute Value (SMAV)

$$SMAV_i = \frac{MAV_i}{MMAV}$$

where

$$MMAV = \frac{\sum_{i=1}^{N} MAV_i}{N} \tag{9}$$

and i is the channel and N is the number of sensors.

Co-contraction Index (CCI)

$$CCI = \frac{\frac{1}{N}\sum_{n=1}^{N} X_e}{\frac{1}{N}\sum_{n=1}^{N} X_f} \tag{10}$$

where Xe is activity from the channel designated as the extensor sensor, and X_f is activity from the channel designated as the flexor sensor.

Scaled Co-contraction Index (SCCI)

$$SCCI = \frac{\frac{1}{N}\sum_{n=1}^{N} SMAV_e}{\frac{1}{N}\sum_{n=1}^{N} SMAV_f} \tag{11}$$

where $SMAVe$ is activity from the channel designated as the extensor sensor, and $SMAV_f$ is activity from the channel designated as the flexor sensor.

C. Movement Variability Group

Circular mean (CircMean)

$$\bar{r} = \frac{1}{N}\sum_{i} r_i$$

where

$$r_i = \begin{pmatrix} \cos\alpha_i \\ \sin\alpha_i \end{pmatrix} \tag{12}$$

And α_i is a sample of all data at one instant i.

Circular resultant (CircR)

$$R = \|\bar{r}\| \tag{13}$$

Circular skew (CircSkw)

$$b = \frac{1}{N}\sum_{i=1}^{N} sin2(\alpha_i - \bar{\alpha}) \tag{14}$$

Circular kurtosis (CircKrt)

$$k = \frac{1}{N}\sum_{i=1}^{N} cos2(\alpha_i - \bar{\alpha}) \tag{15}$$

Circular standard deviation (CircStd)

$$s_0 = \sqrt{-1\ln R} \tag{16}$$

Circular variance (CircVar)

$$S = 1 - R \tag{17}$$

Resultant acceleration magnitude (Accel_Mag)

$$AM = \frac{1}{N}\sum_{n=1}^{N}\sqrt{a_{xn}^2 + a_{yn}^2 + a_{zn}^2} \quad (18)$$

where a, is the acceleration from each axis x, y, z.

Resultant acceleration variance (Accel_Var)

$$AV = \frac{1}{N-1}\left(\sum_{n=1}^{N}\sqrt{a_{xn}^2 + a_{yn}^2 + a_{zn}^2} - AM\right)^2 \quad (19)$$

Resultant gyroscopic magnitude (Gyro_Mag)

$$GM = \frac{1}{N}\sum_{n=1}^{N}\sqrt{g_{xn}^2 + g_{yn}^2 + g_{zn}^2} \quad (20)$$

where g, is the inertial data from each axis x, y, z.

Resultant gyroscopic variance (Gyro_Var)

$$GV = \frac{1}{N-1}\left(\sum_{n=1}^{N}\sqrt{g_{xn}^2 + g_{yn}^2 + g_{zn}^2} - GM\right)^2 \quad (21)$$

Acknowledgment

The authors are greatly appreciative of the time and input of the clients and families who helped us build and test this system. They also wish to acknowledge the occupational therapists and research administration staff who supported this project at Holland Bloorview Kids Rehabilitation Hospital and Fondation Ellen Poidatz. Special thanks to the head of the DISI, and to the head of the UFR ST, of University of Evry for lending computers. And to the developer of the game, KasSanity Inc. They also wish to acknowledge the data processing and administrative assistance of the Research Manager.

References

[1] P. Rosenbaum and D. Stewart, "The world health organization international classification of functioning, disability, and health: A model to guide clinical thinking, practice and research in the field of cerebral palsy," *Seminars Pediatric Neurol.*, vol. 11, no. 1, pp. 5–10, Mar. 2004, doi: 10.1016/j.spen.2004.01.002.

[2] M. Oskoui, F. Coutinho, J. Dykeman, N. Jetté, and T. Pringsheim, "An update on the prevalence of cerebral palsy: A systematic review and meta-analysis," *Develop. Med. Child Neurol.*, vol. 55, no. 6, pp. 509–519, Jun. 2013, doi: 10.1111/dmcn.12080.

[3] T. D. Sanger *et al.*, "Definition and classification of negative motor signs in childhood," *Pathologica*, vol. 118, no. 5, pp. 2159–2167, Nov. 2006, doi: 10.1542/peds.2005-3016.

[4] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and EMG sensors," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 41, no. 6, pp. 1064–1076, Nov. 2011.

[5] J. Liu, X. Sheng, D. Zhang, N. Jiang, and X. Zhu, "Towards zero retraining for myoelectric control based on common model component analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 4, pp. 444–454, Apr. 2016, doi: 10.1109/TNSRE.2015.2420654.

[6] J. Olias, R. Martin-Clemente, M. A. Sarmiento-Vega, and S. Cruces, "EEG signal processing in MI-BCI applications with improved covariance matrix estimators," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 5, pp. 895–904, May 2019, doi: 10.1109/TNSRE.2019.2905894.

[7] M. Rassafiani and R. Sahaf, "Hypertonicity in children with cerebral palsy: A new perspective," *Iranian Rehabil. J.*, vol. 9, p. 9, Dec. 2011.

[8] E. Dursun, N. Dursun, and D. Alican, "Effects of biofeedback treatment on gait in children with cerebral palsy," *Disabil. Rehabil. Int. Multidiscip. J*, vol. 26, no. 2, p. 116, 2004.

[9] S. Lobov, V. Mironov, I. Kastalskiy, and V. Kazantsev, "A spiking neural network in sEMG feature extraction," *Sensors*, vol. 15, no. 11, pp. 27894–27904, Nov. 2015.

[10] J. W. Yoo, D. R. Lee, Y. J. Sim, J. H. You, and C. J. Kim, "Effects of innovative virtual reality game and EMG biofeedback on neuromotor control in cerebral palsy," *Biomed. Mater. Eng*, vol. 24, p. 3613, Jan. 2014.

[11] D. C. Rios *et al.*, "NeuroGame therapy to improve wrist control in children with cerebral palsy: A case series," *Develop. Neurorehabilitation*, vol. 16, no. 6, pp. 398–409, Dec. 2013.

[12] A. Macintosh *et al.*, "The design and evaluation of biofeedback in motor therapy gaming," *Assistive Technol.*, vol. 17, pp. 1–9, Apr. 2020.

[13] P. Visconti, F. Gaetani, G. A. Zappatore, and P. Primiceri, "Technical features and functionalities of Myo armband: An overview on related literature and advanced applications of myoelectric armbands mainly focused on arm prostheses," *Int. J. Smart Sens. Intell. Syst.*, vol. 11, no. 1, pp. 1–25, 2018, doi: 10.21307/ijssis-2018-005.

[14] F. Vrins, J. A. Lee, M. Verleysen, V. Vigneron, and C. Jutten, "Improving independent component analysis performances by variable selection," in *Proc. IEEE 13th Workshop Neural Netw. Signal Process.*, Sep. 2003, pp. 359–368, doi: 10.1109/NNSP.2003.1318035.

[15] N. Meinshausen, "Quantile regression forests," *J. Mach. Learn. Res.*, vol. 7, pp. 983–999, Jun. 2006.

[16] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[17] Mathworks. *Bag of Decision Trees—MATLAB—MathWorks Benelux*. Accessed: Sep. 24, 2019. [Online]. Available: https://nl.mathworks.com/help/stats/treebagger-class.html

[18] A. Phinyomark, S. Hirunviriya, C. Limsakul, and P. Phukpattaranont, "Evaluation of EMG feature extraction for hand movement recognition based on Euclidean distance and standard deviation," in *Proc. Electr. Eng./Electron. Comput. Telecommun. Inf. Technol. (ECTI-CON), Int. Conf.*, May 2010, pp. 856–860.

[19] A. A. Adewuyi, L. J. Hargrove, and T. A. Kuiken, "Evaluating EMG feature and classifier selection for application to partial-hand prosthesis control," *Frontiers Neurorobotics*, vol. 10, p. 15, Oct. 2016, doi: 10.3389/fnbot.2016.00015.

[20] S. Arjunan and D. Kumar, "Decoding subtle forearm flexions using fractal features of surface electromyogram from single and multiple sensors," *J. NeuroEng. Rehabil.*, vol. 7, no. 1, p. 53, 2010, doi: 10.1186/1743-0003-7-53.

[21] I. M. Donovan, J. Puchin, K. Okada, and X. Zhang, "Simple space-domain features for low-resolution sEMG pattern recognition," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 62–65, doi: 10.1109/EMBC.2017.8036763.

[22] P. Berens, "CircStat: AMATLABToolbox for circular statistics," *J. Stat. Softw.*, vol. 31, no. 10, pp. 1–21, 2009, doi: 10.18637/jss.v031.i10.

[23] Mathworks. *Fit Multiclass Models for Support Vector Machines or Other Classifiers—MATLAB Fitcecoc Mathworks Benelux*. Accessed: Sep. 24, 2019. [Online]. Available: https://nl.mathworks.com/help/stats/fitcecoc.html

[24] S. Escalera, O. Pujol, and P. Radeva, "Separability of ternary codes for sparse designs of error-correcting output codes," *Pattern Recognit. Lett.*, vol. 30, no. 3, pp. 285–297, Feb. 2009, doi: 10.1016/j.patrec.2008.10.002.

[25] T. D. Sanger, "Bayesian filtering of myoelectric signals," *J. Neurophysiol.*, vol. 97, no. 2, pp. 1839–1845, Feb. 2007.

[26] A.-C. Eliasson *et al.*, "The manual ability classification system (MACS) for children with cerebral palsy: Scale development and evidence of validity and reliability," *Develop. Med. Child Neurol.*, vol. 48, no. 7, pp. 549–554, Feb. 2007.

[27] A. MacIntosh, E. Desailly, N. Vignais, V. Vigneron, and E. Biddiss, "A biofeedback-enhanced therapeutic exercise video game intervention for young people with cerebral palsy: A randomized single-case experimental design feasibility study," *PLoS ONE*, vol. 15, no. 6, Jun. 2020, Art. no. e0234767, doi: 10.1371/journal.pone.0234767.

[28] K. L. Gwet, *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters?; [A Handbook for Researchers, Practitioners, Teachers & Students*], 3th ed. Gaithersburg, MD, USA: Advanced Analytics, 2012.

[29] K. L. Gwet, "Computing inter-rater reliability and its variance in the presence of high agreement," *Brit. J. Math. Stat. Psychol.*, vol. 61, no. 1, pp. 29–48, May 2008, doi: 10.1348/000711006x126600.

[30] M. Bekkar, H. K. Djemaa, and T. A. Alitouche, "Evaluation measures for models assessment over imbalanced data sets," *J. Inf. Eng. Appl.*, vol. 3, no. 10, 2013.

[31] MATLAB. *Choose Classifier Options—MATLAB & Simulink - MathWorks Benelux*. Accessed: Oct. 1, 2019. [Online]. Available: https://nl.mathworks.com/help/stats/choose-a-classifier.html

[32] D. K. H. Lim, B. McFee, and G. Lanckriet, "Robust Structural Metric Learning," in *Proc. 30 th Int. Conf. Mach. Learn.*, vol. 28, 2013, p. 9.

[33] A. Field, *Discovering Statistics Using IBM SPSS Statistics*. Newbury Park, CA, USA: Sage, 2013.

[34] X. Li, P. Fang, L. Tian, and G. Li, "Increasing the robustness against force variation in EMG motion classification by common spatial patterns," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 406–409.

[35] X. Li *et al.*, "Towards reducing the impacts of unwanted movements on identification of motion intentions," *J. Electromyogr. Kinesiol.*, vol. 28, pp. 90–98, Jun. 2016.

[36] E. J. Scheme, K. B. Englehart, and B. S. Hudgins, "Selective classification for improved robustness of myoelectric control under nonideal conditions," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1698–1705, Jun. 2011.

[37] Q. Ding, X. Zhao, J. Han, C. Bu, and C. Wu, "Adaptive hybrid classifier for myoelectric pattern recognition against the interferences of outlier motion, muscle fatigue, and electrode doffing," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 5, pp. 1071–1080, May 2019, doi: 10.1109/TNSRE.2019.2911316.

[38] F. Duan and L. Dai, "Recognizing the gradual changes in sEMG characteristics based on incremental learning of wavelet neural network ensemble," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 4276–4286, May 2017.

[39] X. Chen, D. Zhang, and X. Zhu, "Application of a self-enhancing classification method to electromyography pattern recognition for multifunctional prosthesis control," *J. NeuroEng. Rehabil.*, vol. 10, no. 1, p. 44, 2013.

[40] Q. Huang, D. Yang, L. Jiang, H. Zhang, H. Liu, and K. Kotani, "A novel unsupervised adaptive learning method for long-term electromyography (EMG) pattern recognition," *Sensors*, vol. 17, no. 6, p. 1370, Jun. 2017.

[41] M. E. Benalcazar, A. G. Jaramillo, Jonathan, A. Zea, A. Paez, and V. H. Andaluz, "Hand gesture recognition using machine learning and the myo armband," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 1040–1044.

[42] I. Mendez *et al.*, "Evaluation of the myo armband for the classification of hand motions," in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Jul. 2017, pp. 1211–1214.

[43] N. J. O'Dwyer, L. Ada, and P. D. Neilson, "Spasticity and muscle contracture following stroke," *Brain*, vol. 119, no. 5, pp. 1737–1749, 1996.

[44] N. H. Mayer, "Clinicophysiologic concepts of spasticity and motor dysfunction in adults with an upper motoneuron lesion," *Muscle Nerve*, vol. 20, no. S6, pp. 1–14, 1997.

[45] G. Sheean, "The pathophysiology of spasticity," *Eur. J. Neurol.*, vol. 9, pp. 3–9, May 2002.

[46] P. Kieliba, P. Tropea, E. Pirondini, M. Coscia, S. Micera, and F. Artoni, "How are muscle synergies affected by electromyography pre-processing?" *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 4, pp. 882–893, Apr. 2018, doi: 10.1109/TNSRE.2018.2810859.

[47] A. d'Avella, L. Fernandez, A. Portone, and F. Lacquaniti, "Modulation of phasic and tonic muscle synergies with reaching direction and speed," *J. Neurophysiol.*, vol. 100, no. 3, pp. 1433–1454, Sep. 2008.

[48] J. Roh, W. Z. Rymer, E. J. Perreault, S. B. Yoo, and R. F. Beer, "Alterations in upper limb muscle synergy structure in chronic stroke survivors," *J. Neurophysiol.*, vol. 109, no. 3, pp. 768–781, Feb. 2013.

[49] P. Tropea, V. Monaco, M. Coscia, F. Posteraro, and S. Micera, "Effects of early and intensive neuro-rehabilitative treatment on muscle synergies in acute post-stroke patients: A pilot study," *J. NeuroEng. Rehabil.*, vol. 10, no. 1, p. 103, 2013.

[50] W. Yang, K. Wang, and W. Zuo, "Neighborhood component feature selection for high-dimensional data," *J. Comput.*, vol. 7, no. 1, pp. 161–168, Jan. 2012, doi: 10.4304/jcp.7.1.161-168.