Human-in-the-Loop Data Acquisition Improves Myoelectric Control of a Prosthetic Hand

Andrea Gigli[1], **Donato Brusamento**[1], **Roberto Meattini**[2], **Claudio Melchiorri**[2], **and Claudio Castellini**[1]

[1]Institute of Robotics and Mechatronics, German Aerospace Center (DLR), Wessling, Germany

[2]Department of Electrical, Electronic and Information Engineering (DEI), University of Bologna, Bologna, Italy

E-mail: `andrea.gigli@dlr.de`

August, 2020

Abstract.

Myocontrol is unreliable, which hinders its deployment in the clinical practice and in daily-living activities, and keeps the acceptance level of dexterous prosthetic devices low. One cause for this is the poor generalization of the underlying machine-learning models to untrained conditions. This may depend on the standard data acquisition method, consisting of collecting training data in an open loop, non-interactively. The problem could be reduced by giving the user an active role in the training of the prosthesis. This is an emerging trend in myocontrol of upper-limb self-powered prostheses: that it should be interactive, involving the user since the beginning and during the whole usage. In this study, 18 non-disabled participants compared one open-loop and two human-in-the-loop multi-arm-position data acquisition protocols. As opposed to open-loop, during human-in-the-loop acquisition an acoustic signal urged the participant to hover with the arm in specific regions of her peri-personal space, de facto acquiring more data where needed, possibly further mending the limb-position effect. The three protocols were compared on daily-living-like manipulation tasks performed with a prosthetic hand. Our results confirm that human-in-the-loop acquisition outperforms open-loop acquisition both objectively and subjectively, suggesting that interaction is fundamental to improve myocontrol.

Keywords: human-prosthesis interaction, myoelectric control, prosthetic hand, pattern recognition

Submitted to: *J. Neural Eng.*

1. Introduction

The loss of an upper limb can affect the ability to carry out essential activities of daily living (ADLs) [1]. Even though modern myoelectric prostheses offer the possibility to restore some of the lost functionalities, their adoption in clinical applications remains limited. High rejection rates have been documented for powered upper-limb prostheses in favor of using passive devices or, even, no device at all [2]. Among the reasons for this are the high cost of myoelectric devices and the unintuitiveness and unreliability of the control during ADLs [3]. Simultaneous and proportional (s/p) myocontrol has been investigated in recent years to favor a natural interaction with prosthetic hands [4]. In s/p myocontrol, a regression-based model is used to map voluntary muscular contractions into independent motor commands for the prosthesis' degrees of freedom (DoFs). The model is typically trained with labeled myoelectric signals recorded at the user's forearm via surface electromyography (sEMG).

Such approaches are capable of producing models with considerable predictive power, provided that the control signal's characteristics remain similar between training and testing conditions. However, sEMG changes due to variations of skin conductivity, electrode placement and limb position, as well as to fatigue phenomena and the evolution of the user's cognitive capabilities [5].

We hereby focus on the limb-position effect [6, 7], whose negative impact on online myocontrol performance has been characterized in multiple studies [8, 9]. Proposed solutions to this problem include computing limb-position-invariant features from the myoelectric signal [10], using prediction models that are resistant to the noise produced by the arm movement [11], or acquiring training data in multiple arm positions. Static and dynamic multi-position acquisition protocols have been proposed, the first involving the repetition of target hand gestures in multiple arm configurations [6], the latter requiring the execution of predefined arm movements [12, 13]. Both approaches proved to enhance myocontrol performance compared to single-position training, and dynamic training also allowed reducing the duration and the physical effort required to cover the space of possible arm configurations [14].

To the best of our knowledge, however, all multi-position acquisition protocols found in the literature do not actively involve the user, who traditionally performs a predefined sequence of arm configurations or movements without knowing to which extent each of them contributes to the improvement of the myocontrol model. As opposed to this, we propose that the user should be notified in which limb positions more training data is needed, since such positions are more problematic for the model. Therefore, we introduce the concept of *user-driven data acquisition* as a way to generate s/p myocontrol models that are robust to the limb position effect. Closing the myocontrol loop at some point of the human-robot interfacing chain involves using different sensory modalities to feed the user information related to any step of the myocontrol process [15, 16].

This idea can be seen as a special case of the *human-in-the-loop* paradigm [17], in which bidirectional user-prosthesis interaction is enforced. Importantly, this approach can be seamlessly integrated within the framework of interactive myocontrol [18, 19, 20], where man-machine bilateral adaptation is put in place exploiting incremental learning myocontrol. Sources of information in myoelectric control include biometric signals measured from the user (biofeedback) [21], model's predictions [22], and interaction forces between the prosthesis and the touched objects [23]. The use of feedback proved beneficial to restore the sense of touch and proprioception in impaired limbs [16], to improve realtime regulation of the grasp force, to help the user develop more effective cognitive models of the prosthesis controller [24], and to enhance the sense of trust and embodiment in the prosthesis [25]. In general, human-in-the-loop myocontrol has been employed to improve the usability of the prosthesis at runtime, without directly affecting the underlying control model. Feedback was first utilized during data acquisition to guide the model building process in the work of Hahne et al. [22]. They designed a human-in-the-loop data acquisition strategy in which the user was encouraged to produce progressively better muscle contraction patterns prompted by instantaneous feedback about the model's improvement due to the generated patterns. Their results showed that human-in-the-loop model training is a key factor for co-adaptive prosthetic control, allowing both the user and the device to converge to a synergistic control strategy.

The user-driven data acquisition paradigm, here introduced, aims at improving the robustness of the myocontrol model with respect to the limb position effect. It integrates simultaneous dynamic data acquisition and incremental model building with a feedback signal related to the usefulness of the recorded data. The feedback is designed to help the user identify and collect training data in the areas of the input space, i.e., arm configurations, where the model prediction is inaccurate. We compare one standard, "user-blind" acquisition protocol and two variants of the novel user-driven data acquisition protocols, all based on the dynamic acquisition presented in Gigli et al. [14]. Both user-driven procedures adopt the same feedback mechanism, but one of them also integrates an automatic sample selection criterion to discard unnecessary training samples and reduce the number of model updates.

2. Materials and Methods

This study evaluates the effects of using a feedback signal to guide the acquisition of training data for myoelectric controllers of prosthetic hands. The performances of one standard open-loop and two human-in-the-loop data acquisition procedures were compared based on the controllability of a prosthetic hand during a series of realistic manipulation tasks.

2.1. Participants

Eighteen able-bodied persons (aged 26.3 ± 4.6 years, 16 men and 2 women) participated in the experiment. Twelve participants had no prior experience in myoelectric control, while six had already used myoelectric prosthetic hands in previous user studies. Every participant received an oral and written description of the experiment and signed an informed consent form. The study was conducted at the German Aerospace Center (DLR) according to the WMA Declaration of Helsinki and approved by DLR's internal committee for personal data protection.

2.2. Experimental Setup

The muscular activity of the forearm of the dominant arm was measured using a Myo armband‡ by Thalmic Labs placed about 5 cm below the elbow. The bracelet comprised eight sensors, each recording an sEMG signal at a sampling rate of 200 Hz. A standard quick-release prosthetic connector fixed to a wrist/hand orthotic splint made it possible to anchor the prosthesis at the extremity of sound limbs. An i-LIMB Ultra Revolution prosthetic hand§ by Touch Bionics (now Össur) allowed independent flexion/extension of the five fingers and abduction/adduction of the thumb through six motors under direct current control. The devices communicated via a serial-port-over-Bluetooth with a laptop used to run the myocontrol software. The acoustic feedback was reproduced using the speakers of the laptop. A custom software suite written in the C# language provided the graphical interface to coordinate the data acquisition, labeled and processed sEMG data, generated the feedback signal, and implemented realtime myocontrol.

The experiment took place in a domestic-like laboratory environment. We arranged several household objects on a table, two shelves, and on the floor. We placed the table 40 cm next to the shelves, and we regulated its height to match the waist level of each participant. The shelves were 40 cm and 150 cm high. The study was videotaped in order to measure the participant's performance after the experiment. Figure 1A shows the experimental setup.

2.3. Incremental model building

The 8-channels sEMG readings were preprocessed in realtime upon collection. The measurement from each channel was rectified, computing its absolute value, and low-pass filtered using a second-order Butterworth filter with a cutoff frequency of 1 Hz.

The data acquisition software labeled incoming training samples with the activation commands for the motors of the prosthetic hand's fingers. Each command consisted of a normalized velocity ranging between 0 and 1, corresponding to extending or flexing the finger with maximum speed. Since all the hand gestures considered in this experiment could be realized by controlling one subset of the fingers with the same velocity command, the model had effectively 3 DoFs.

Training data was collected only for extreme velocity commands values, that is, for hand gestures in which each finger was either fully extended or fully flexed. Intermediate velocity command values were excluded because they could lead to inaccuracies in the recorded data due to the participants' different reaction times [26]. Previous works, such as [27, 14], showed that regression models resulting from this training procedure still yield effective s/p control.

To provide appropriate feedback guidance during the data acquisition, it was necessary to incorporate each training sample into the model quickly upon collection. Therefore, we trained the s/p control model using an instance of incremental ridge regression (iRR) with random Fourier features (RFFs). iRR builds a regression model incrementally by computing rank-one model updates when new training data is available. The iRR formulation allowed us to update the model and generate predictions with bounded time and space complexity. RFFs is a nonlinear mapping of the input space into a high-dimensional feature space obtained by using sinusoidal basis functions that have randomly sampled frequencies. By drawing those frequencies from an adequate probability distribution and choosing a sufficiently high mapping dimensionality, iRR with RFF approximates ridge regression with a Gaussian kernel [27]. Consequently, RFFs extend the capacity of iRR to perform nonlinear regression while maintaining the properties of incrementality and boundedness of the model update. This is relevant in applications that require online learning of nonlinear regression models [27, 28, 29, 14]. A detailed description of iRR-RFF and its use for s/p myoelectric control can be found in [27].

‡ `https://support.getmyo.com/hc/en-us/articles/203398347-Getting-started-with-your-Myo-armband`
§ `https://www.ossur.com/en-us/prosthetics/arms/i-limb-ultra`

The prediction function of iRR-RFF is

$$\hat{\mathbf{y}} = \mathbf{W} \cdot \Phi(\mathbf{x}) \quad (1)$$

where $\mathbf{x} \in \mathbb{R}^d$ is an input sample, $\Phi : \mathbb{R}^d \to \mathbb{R}^D$ is a nonlinear RFF mapping, $\mathbf{W}$ is an $M \times D$ matrix of scalar weights, and $\hat{\mathbf{y}} \in \mathbb{R}^M$ is the computed prediction. In this experiment, training pairs $\{\mathbf{x}, \mathbf{y}\}$ consisted of an sEMG measurement and the corresponding velocity commands for the fingers' motors. The input and output dimensionality of the model were $d = 8$ and $M = 3$. The regularization parameter λ of the ridge regression was set to 1, while the bandwidth γ and the dimensionality D of the RFF mapping were set to 0.1 and 300, respectively. The model weights were initialized to zero before the data acquisition, $\mathbf{W} = \mathbf{0}_{M,D}$.

2.4. Acoustic feedback and sample selection

The idea behind human-in-the-loop data acquisition is to guide the participant with an appropriate feedback signal in order to maximize the amount of informative and non-redundant data collected in a fixed amount of time. In an online learning problem, the informativeness of a correctly-labeled training sample $\{\mathbf{x}, \mathbf{y}\}$ for the model can be evaluated based on the prediction error

$$e_p(\mathbf{y}, \hat{\mathbf{y}}) = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 \quad (2)$$

where $\hat{\mathbf{y}}$ is the prediction of the sample using the model. A small prediction error indicates that the model can accurately predict the label for that sample and, therefore, the sample might be redundant for the model. A significant prediction error, instead, indicates that the model fails to predict the right label and may improve by integrating the training sample. For the sake of clarity, we omit the argument of the prediction error in the remainder of the paper.

2.4.1. Feedback signal For our purposes, we designed an acoustic feedback signal with a fixed tone and variable volume. The volume of the signal ranged between 0 and a maximum value V and varied proportionally with the prediction error according to

$$f(e_p) = \max\left\{0, \min\left\{ae_p^2 + \frac{V - a\theta_u^2}{\theta_u}e_p, V\right\}\right\} \quad (3)$$

in which a was a scalar regulating the quadratic relation between error and volume, and θ_u was a threshold related to the prediction error. We set the values of the parameters to $V = 0.5$, $a = 70$, and $\theta_u = 0.05\sqrt{3}$. The value of θ_u corresponded to 5 % of the maximum theoretical value of the prediction error in our experiment, which was predicting an open hand gesture instead of a power grasp gesture.

(*a*) (*b*)

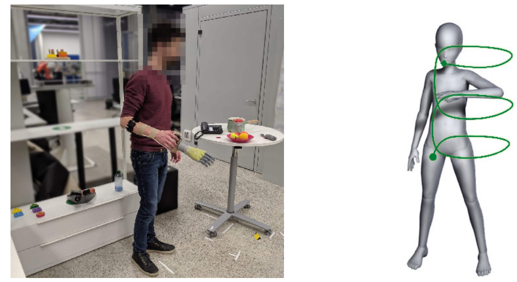


Figure 1. *Experimental setup and arm motion during data acquisition.* *(a)* The prosthetic system comprised a Myo armband by Thalmic Labs for sEMG reading, and an i-LIMB Ultra Revolution prosthetic hand by Össur. The experimental setup included common household objects placed onto one table and two shelves. The speakers of the control laptop provided acoustic feedback. *(b)* Participants wore the prosthetic system throughout the data acquisition. Every data acquisition routine required to perform several target hand gestures while moving the arm in a predefined trajectory. The motion proceeded from the circle to the square with the palm oriented downward and continued in the opposite direction with the palm oriented upward.

2.4.2. Sample Selection We also used the prediction error to discard possibly redundant training samples. We defined a sample selection criterion to update the model only with those training samples for which

$$e_p \geq \theta_u \quad (4)$$

where θ_u is the update threshold defined before.

2.5. Experimental protocol

Every participant tested all the data acquisition strategies. We counterbalanced possible learning effects by administering the strategies to the participants in randomized order. We assigned each of the six permutations of the training conditions to one experienced and two naive participants picked at random. After each data acquisition, the resulting myocontrol model was tested in a sequence of realtime manipulation tasks. Participants repeated the sequence of tasks three times. The first two repetitions of the sequence allowed the participants to familiarize themselves with the prosthetic system and the myocontrol model, while the third one was used to measure the myocontrol performance. For this reason, we referred to the third repetition of the task sequence as a performance evaluation session.

2.5.1. Data acquisition All the data acquisition strategies required the participants to perform several target hand gestures while moving their arm in the reachable space. We selected three target hand gestures: namely a power grasp, a resting hand,

and an index pointing. The selection was based on their relevance in ADLs, according to the literature [30]. The target hand gestures were acquired in the order reported above in every data acquisition. Since the myocontrol model was built incrementally, the use of different orders would have possibly led to incomparable models.

Before the experiment, participants were explained the data acquisition protocols and were asked to practice them. Emphasis was put into enforcing a consistent arm movement across participants and strategies. Meanwhile, the volume of the speakers was regulated so to ensure that the feedback was distinctly audible. Nonetheless, the experimenter supervised the data acquisition and provided direct guidance when the participants performed the arm movement at the wrong pace or ignored the acoustic feedback.

Participants donned the prosthetic system on the dominant arm at the beginning of the experiment, and no adjustment of the sensors was allowed after that. Wearing the prosthesis during the data acquisition reduced the differences between the training and testing conditions caused by factors such as the electrodes' placement and the weight of the prosthetic device.

Open-Loop Data Acquisition (OL-DA) OL-DA adapted the dynamic acquisition presented in our previous work [14] to the setup of this study. Participants performed each target hand gesture while moving their arm in a predefined trajectory. During the procedure, they did not receive any feedback. The model was built incrementally in realtime with each new training sample, as detailed in section 2.3. The trajectory uniformly covered the reachable space of the participant with a helical movement, Figure 1B. The movement was performed with constant speed from the level of the waist to the level of the head with the palm oriented downward; it continued in the opposite direction with the palm oriented upward. This whole sequence was repeated twice, without interruptions. The motion lasted 45 s for each hand gesture and took 135 s in total. The procedure is synthesized in Algorithm 1.

Algorithm 1: Open-Loop Data Acquisition

```
Input: stream of sEMG samples x
init model to zero;
foreach hand gesture g do
    while participant performs g do
        acquire new sample x;
        update model with {x, label(g)};
    end
end
```

Human-in-the-Loop Data Acquisition (HL-DA) HL-DA extended OL-DA with the acoustic feedback detailed in section 2.4.1. The acquisition software used the incoming training samples to generate the acoustic feedback and to build the myocontrol model in realtime. Participants had to perform the desired grasp and follow the usual arm trajectory while modulating the arm's velocity based on the feedback. They should proceed with the same speed used during open-loop acquisition when the feedback was not audible and hover with the arm in the areas where the feedback intensity increased. Since the feedback was proportional to the prediction error, this procedure led the participants to collect more data in critical arm configurations. The model incrementality prevented participants from slowing down indefinitely in critical areas of the reachable space. Training samples were continuously integrated into the myocontrol model, which immediately reduced the prediction error and, consequently, the volume of the feedback signal. The acquisition of each gesture lasted 45s. Differently from OL-DA, however, participants were not expected to cover the whole trajectory twice per gesture. The procedure is synthesized in Algorithm 2.

Algorithm 2: Human-in-the-Loop Data Acquisition

```
Input: stream of sEMG samples x
init model to zero;
foreach hand gesture g do
    while participant performs g do
        acquire new sample x;
        compute prediction error;
        generate acoustic feedback;
        update model with {x, label(g)};
    end
end
```

Human-in-the-Loop Data Acquisition with Sample Selection (HLSS-DA) HLSS-DA was obtained by integrating HL-DA with the sample selection criterion described in section 2.4.2. All the incoming training samples were used to generate the acoustic feedback, but only a limited number of non-redundant samples were selected and used to build the myocontrol model in realtime. Participants perceived no formal difference between the two human-in-the-loop acquisition routines. The procedure is synthesized in Algorithm 3.

2.5.2. Realistic myocontrol tasks After every data acquisition, the resulting myocontrol model was tested by engaging the participants in a series of five

Algorithm 3: Human-in-the-Loop Data Acquisition with Sample Selection

```
Input: stream of sEMG samples x
init model to zero;
foreach hand gesture g do
    while participant performs g do
        acquire new sample x;
        compute prediction error;
        generate acoustic feedback;
        if predictionerror > threshold then
            update model with {x, label(g)};
        end
    end
end
```

manipulation tasks. The tasks were inspired by realistic ADLs proposed in assessment protocols for prosthetic control, such as ACMC [31] and SHAP [32]. The tasks are described in Table 1. In the case of bimanual tasks, we assigned each action of the task either to the prosthetic hand or the sound hand.

The experimental protocol required the completion of all the tasks. If an object was dropped during a manipulation task, the experimenter brought it back to the place where it had been grasped, and the task continued from where it failed. The time needed to reset the position of the object was excluded from the final evaluation of performance. If repeated instabilities of the prosthesis hindered the execution of one task, the experimenter or the participant could suspend the task and request an additional model update. On-demand model updates were obtained with shorter versions of the data acquisition procedure performed at the beginning of the corresponding experimental session. Participants were instructed to hold the malfunctioning hand gesture while randomly moving the arm for 10 s in the area where the task failed, possibly enforcing movements of the shoulder, elbow, and forearm.

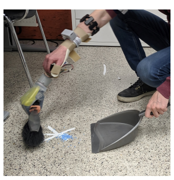
2.6. Performance evaluation

We evaluated the effectiveness of each data acquisition procedure based on the duration of the third repetition of the task sequence. We chose the task execution time as an objective measure of myocontrol performance because it is at the base of many clinical assessment protocols for the hand function [33], it is simpler to measure than other metrics, and does not require a trained examiner. Subjective performance measures were obtained by asking the participants to rate the controllability of the prosthetic system and the difficulty of the tasks in a questionnaire at the end of the experiment. The controllability of the prosthetic system resulting from each training condition was reported on a visual analog scale (VAS) ranging from "very easy to control" to "very difficult to control". Similarly, each task's difficulty was quantified on a VAS ranging from "very difficult" to "very easy". We verified if any of the acquisition strategies resulted in better controllability or faster task execution compared to the others, which could indicate a more robust myocontrol model and, therefore, better training data. A Shapiro-Wilk test revealed that the task duration and the results of the questionnaire were not normally distributed across participants. For this reason, we used a Freedman test to identify differences in the average value of the statistics of the three training conditions. When the test indicated significant differences, we used repeated post-hoc Wilcoxon signed-rank tests to compare pairs of conditions. We set the significance level of all the tests to $\alpha = 0.05$, and we controlled the inflation of the significance level during repeated pairwise tests by operating a Bonferroni adjustment of the p-value [34]. In this paper, we reported unadjusted p-values (p) for the Friedman tests and Bonferroni-adjusted p-values ($\hat{p}$) for the post-hoc pairwise tests.

3. Results

The performance of the myocontrol model was measured by the duration of the tasks during the third repetition of the task sequence, i.e., the performance evaluation session. Figure 2A reports the duration of the evaluation session corresponding to the three data acquisition strategies. A Friedman test, followed by pairwise post-hoc Wilcoxon tests, revealed that the evaluation session in the HL-DA condition was significantly faster than in the OL-DA condition (average tasks sequence duration of 166.0 s versus 198 s, $W = 19.5$, $\hat{p} = 0.012$). The average duration of the task sequence in the HLSS-DA condition, 183 s, did not differ significantly from those of the other conditions. Figure 2B and Figure 2C report the performance of naive and experienced participants. For every training strategy, experienced participants completed the evaluation session faster than naive participants. Although not supported by statistical evidence, both groups seemed to perform better after HL-DA compared to OL-DA. The use of feedback during data acquisition reduced the average duration of the performance evaluation session by 15 % for naive participants and by 19 % for experienced participants. For both groups, the mean duration of the tasks after HLSS-DA was characterized by high variability, and its average value was between those of the other two conditions. Figure 2D describes the performance of the participants during the individual tasks. Friedman tests were performed for each task and confirmed

Table 1. Detailed description of the tasks in each performance evaluation repetition.

Task	**Name**	**Description**
	Pour water	A bottle and a jar are placed, respectively, on the lower shelf and on the table. Grasp the bottle[p], unscrew the cap[s], place the bottle and the cap on the table. Take the jar[s], unscrew the lid[p], and put it on the table. Take the bottle[p] and pour the content into the jar. Close the jar[p] and put it on the table. Take the bottle[p], close it[s], and bring it back to the lower shelf.
	Serve food	A pot, a plate containing three tennis balls, and a spoon are laid on the table. Use the spoon[p] to bring the balls from the plate to the pan. Grab the pot by the handle łand tilt it by about 80 degrees, scoop the balls from the pot to the plate using the spoon[p].
	Phone and rolling ball	A telephone is on the table. Dial[p] a sequence of numbers on the phone (1 to 9, 9 to 1, 0, "dial") with an index pointing gesture. A small ball is on the floor, and a target position is marked on the floor about one meter away. Use the index pointing gesture to push the ball[p] to the target position.
	Pegboard	Three wooden shapes from one pegboard game are laid on the lower shelf, while the base is laid on the higher shelf. Pick[p] each shape and stack it to the corresponding peg.
	Sweep the floor	A hand broom and a dustpan are placed on the lower shelf, while a bowl and some gravels are laid on the floor. Grab hand broom[p] and dustpan[s], sweep the gravels onto the dustpan, empty the dustpan in the bowl, and bring the hand broom and the dustpan back to the lower shelf.

[p] prosthetic hand; [s] sound hand.

significant differences in completion time for the third task ($\chi^2(2) = 7.4$, $p = 0.024$). Post-hoc tests, however, failed to identify differences between any pair of conditions, which could be caused by the application of a conservative Bonferroni adjustment to the p-value. Nonetheless, the average duration of every task after HL-DA was slightly lower than after OL-DA. The performances of HLSS-DA remained equivalent to those of the other strategies.

Figure 3A and Figure 3B show the duration of the three repetitions of the task sequence for naive and experienced participants. We referred to these repetitions as the first and second familiarization sessions (F1 and F2), and the performance evaluation session (E). The participants tested the data collection strategies in randomized orders so to counterbalance possible transfer learning effects. Therefore, the results displayed in the figure follow a chronological order within each training condition but not across different conditions. In all the training conditions, naive participants completed the performance evaluation session around 24 % faster than the first familiarization session. Friedman tests confirmed that the reduction of the task completion time during the three repetitions was significant in every training condition ($\chi^2(2) = 9.5$, $p = 0.009$ for OL-DA; $\chi^2(2) = 18.2$, $p < 0.001$ for HLSS-DA; $\chi^2(2) = 18$, $p < 0.001$ for HLSS-DA). At the same time, the variability of the results reduced during the familiarization process. The IQR shrunk from 208-302.5 s to 181.5-229 s for OL-DA, from 191.8-282 s to 163.8-197.5 s for HL-DA, and from 152-327.3 s to 140.3-213.5 s for HLSS-DA. These results, taken together, indicate that a strong learning effect took place for naive participants during

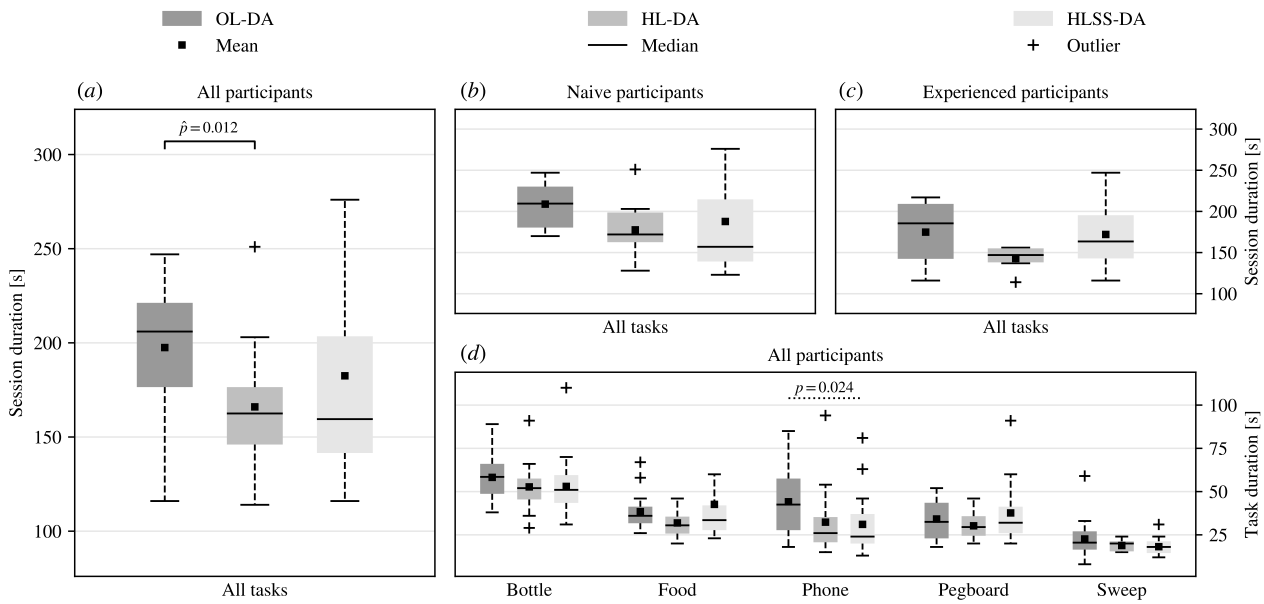


Figure 2. *Duration of the tasks in the performance evaluation session. (a)* Participants completed the evaluation session significantly faster in the HL-DA condition compared to the OL-DA condition ($\hat{p}$ Bonferroni-adjusted). The performance in the HLSS-DA condition did not differ significantly from those of the other conditions. The same could be observed by either considering the naive *(b)* or the experienced *(c)* participants. *(d)* The average duration of each task in the HL-DA condition was slightly lower than that measured after OL-DA, although multiple Friedman tests identified significant differences (p unadjusted) only in the duration of the third task (dialing a phone number). In the paper, boxplots' whiskers extend to the most extreme samples within the first quartile -1.5 IQR and the third quartile $+1.5$ IQR.

Table 2. Median amount of training samples acquired and used to build the myocontrol model

Acquisition protocol	# training samples
OL-DA	27284.5 (IQR 25993-30696)^{a,b}
HL-DA	28439.5 (IQR 25620-30440)^{a,b}
HLSS-DA	26879 (IQR 26654-32696)^{a} 7228.5 (IQR 5345-8646)^{b}

^{a}acquired; ^{b}used.

the familiarization process of each strategy. This learning trend was not as evident among the six experienced participants. For them, the reduction of the task sequence duration due to familiarization was supported by statistical evidence only for the HLSS-DA condition ($\chi^2(2) = 9.3$, $p = 0.009$). Nonetheless, the task execution time reduced by approximately 19 % during the familiarization process for all the training strategies.

Table 2 details the median number of training samples acquired by each strategy and the number of samples that were selected to train the myocontrol model. The number of training samples comprised the data acquired during the initial acquisition and during all the on-demand model updates requested by the participants. All the strategies acquired a comparable amount of training samples, about 28000, although with some variations. The median number of acquired training samples was approximately 27000 for OL-DA and HLSS-DA, and 28500 for HL-DA. The median number of on-demand model updates was equal to 0.5 (IQR 0-2) for OL-DA, 1 (IQR 0-2) for HL-DA, and 0 (IQR 0-3) for HLSS-DA. While OL-DA and HL-DA used all the training samples to build the myocontrol mode, HLSS-DA only employed a median of ≈ 7000 samples, roughly corresponding to a quarter of the acquired data.

Figure 4 shows the perceived difficulty of the myocontrol tasks, assessed by the participants in the final questionnaire, and converted into a percentage from 0 % ("very easy") to 100 % ("very difficult"). The ratings seemed to split tasks into two groups. The relative difficulty of pouring water, serving food, and sweeping the floor was relatively low, around 20 % on average. Precision tasks such as dialing a phone number and completing a pegboard were given a higher average difficulty, around 40 %. In particular, a quarter of the participants found it extremely difficult to dial phone numbers, as they reported a difficulty level higher than 75 %, more than what was reported

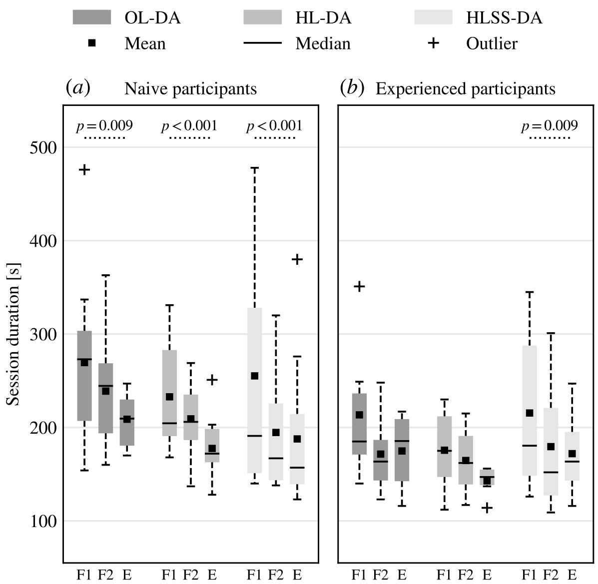


Figure 3. *Effect of learning on the duration of the task sequence.* The three repetitions of the task sequence were labeled F1, first familiarization, F2, second familiarization, and E, performance evaluation session. *(a)* Naive participants showed a significant reduction in the average task completion time due to familiarization (p unadjusted). *(b)* For the experienced participants, the familiarization with the system significantly reduced the duration of the task session only in the HLSS-DA condition.

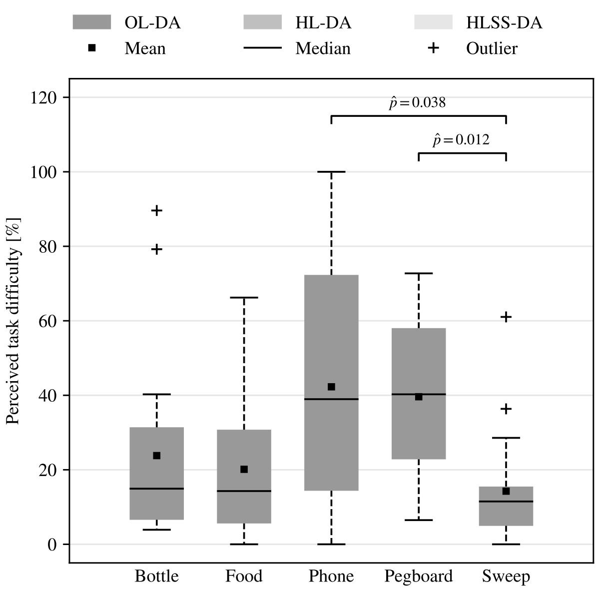


Figure 4. *Perceived tasks difficulty.* On average, participants found more difficult those tasks that required to manipulate small objects (completing the pegboard) or to precisely touch small target areas (dialing a phone number). This result was only partially supported by statistical evidence ($\hat{p}$ Bonferroni-adjusted).

for all the other tasks. A Friedman test confirmed the existence of relevant differences in the perceived complexity of the tasks ($\chi^2(4) = 23.1$, $p < 0.001$). Pairwise post-hoc tests, however, only confirmed that the sweeping task was easier than the dialing task ($W = 9$, $\hat{p} = 0.004$) and the pegboard task ($W = 11$, $\hat{p} = 0.0012$).

The controllability of the prosthetic hand during the myocontrol tasks, reported by the participants in the questionnaire, was converted into a percentage from 0 % ("very difficult to control") to 100 % ("very easy to control"). Overall, the use of feedback during the data acquisition resulted in an improvement of the controllability level of about 10% compared to standard open-loop data acquisition (controllability level of 60 % for OL-DA, 70 % for HL-DA, 71 % for HLSS-DA), Figure 5A. A Friedman test, however, did not support this finding with statistical evidence ($\chi^2(2) = 51$, $p = 0.19$). Naive participants reported lower controllability for every training condition, by about 22% on average, compared to experienced participants. In any training condition, the average controllability reported by naive participants was about 22 % lower than that reported by experienced participants. The ratings of naive participants were mixed. Although the controllability level was slightly higher for the human-in-the-loop acquisition strategies (controllability level of 54 % for the open-loop strategy and 62 % for the human-in-the-loop strategies), the spread of the ratings was exceptionally high, especially for OL-DA (interquartile range, IQR, equal to 28-71 %). Experienced participants, conversely, reported sharper improvements in controllability by following human-in-the-loop training strategies. The average controllability increased from 74 % for OL-DA to 84 % for HL-DA, and 87 % for HLSS-DA. The spread of these results was lower than that observed in naive participants (IQR equal to 56-88 % for OL-DA, 73-94 % for HL-DA, and 78-95 % for HLSS-DA). However, this result was not supported by statistical evidence, possibly due to the limited amount of experienced participants.

4. Discussion

We implemented a human-in-the-loop dynamic data acquisition protocol that used acoustic feedback to induce the user to hover with the arm in the areas of the peri-personal space characterized by poor intent detection, i.e., a discrepancy between the prediction and the ground truth. In the experiment, we have compared three data acquisition strategies for myocontrol, two of which put the participant in the data acquisition loop, as opposed to the third, a traditional one, which acquires data from the

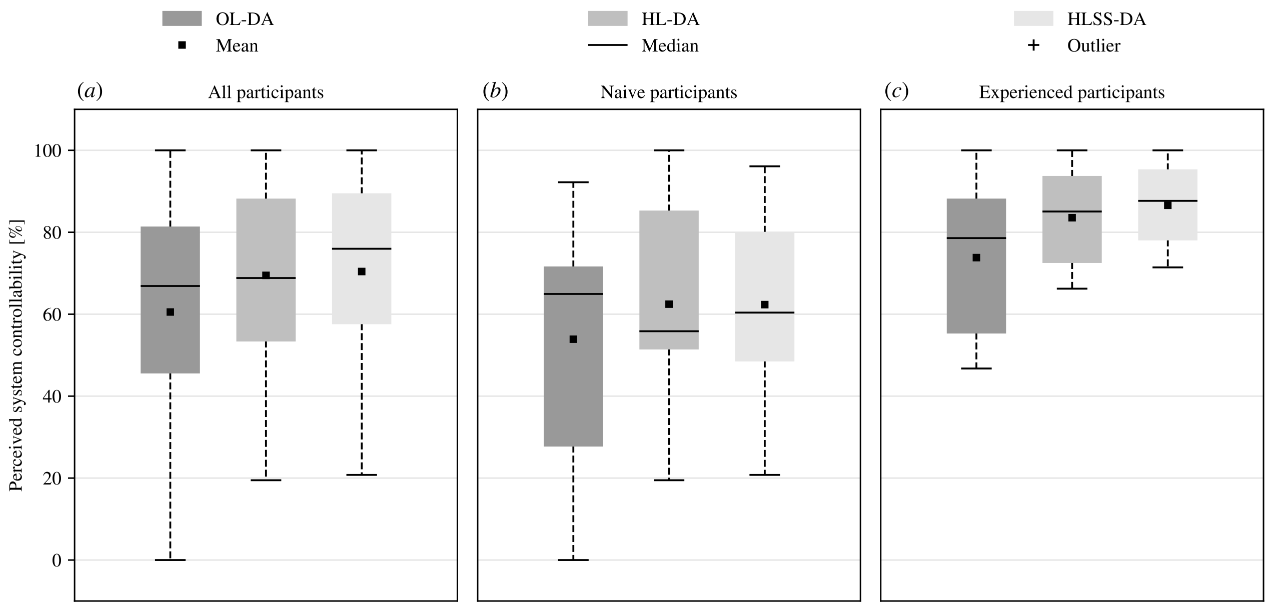


Figure 5. *Perceived controllability of the prosthetic hand. (a)* On average, participants found that the prosthetic system was easier to control after each of the HL data acquisitions compared to the OL data acquisition. *(b)* The ratings reported by the twelve naive participants were mixed and, on average, lower than those of the experienced participants. This caused the high variability observed in the overall results and possibly explained the lack of statistical significance. *(c)* Experienced participants consistently reported that data acquisition routines with feedback resulted in better controllability of the prosthetic system.

participant in a non-interactive way. Our results confirm that *involving the user in the data acquisition procedure yields better myocontrol*, both *objectively*, enabling faster completion of tasks and requiring less computation space and power, and *subjectively*, judged more controllable by participants through quantitative questionnaires.

Participants completed the sequence of manipulation tasks significantly faster when using HL-DA compared to OL-DA, Figure 2. The performance offered by HLSS-DA were characterized by higher variability and did not differ significantly from those of the other acquisition strategies. Even though half of the participants performed equivalently well with HLSS-DA and HL-DA, the other half showed considerably worse performance for HLSS-DA. This could be due to the fact that a model trained with fewer data can be prone to higher instability in the prediction. Possibly, the sample selection criterion has been too strict (with HLSS-DA, the system discarded all samples that determined a prediction error below 5 % of the maximum prediction error). By relaxing that criterion, the results should tend to those of HL-DA, therefore, at least, reducing the variability. However, HLSS-DA allowed for a considerable reduction in the number of samples used to train the machine, of about three-quarters of the total on average. This is especially relevant for realtime applications where the myocontrol model needs to be updated incrementally and therefore repeated model updates are requested for batches of incoming training samples. We then conclude that HLSS-DA can be used as a second choice over HL-DA, only when less computational space is available to the machine learning system. Considering panels B and C of the same figure, it is apparent that human-in-the-loop acquisition improves myocontrol performance for both experienced and naive participants (after a short familiarization phase), which might denote that even experienced myocontrol users might benefit from this system to identify critical areas of the input space. Figure 2D, finally, suggests that the superiority of human-in-the-loop strategies is uniform for each task.

As it was predictable, a quite evident learning effect is present in the performance of all participants, from the two familiarization phases on to the experimental one, Figure 3. Interestingly, this trend characterized naive as well as experienced participants, albeit less so in the latter case. On the one hand, this means that the effect of feedback can be appreciated after a short familiarization with the system (by inexperienced participants, that is). On the other hand, automatic guidance during data collection retains its usefulness over time, since it provides a direct understanding of the state of the myocontrol system at runtime, already during the data acquisition. During the familiarization, participants learn to compensate distracting factors that are inherent in the myocontrol of a prosthetic device, such as the latency and the weight of the hand, and the non-intuitive control of the contraction strength (nonlinear algorithms may not guarantee monotonic mappings of muscle contraction to grip strength). This contributes to reducing the variability of the results and, therefore, helps to observe the effects of interest, e.g., the effect of different data acquisition procedures. During the familiarization process, the performance of naive participants decreased in variability and seemed to tend to those of experienced participants. However, the average duration of the task session at the end of the familiarization process remained slightly higher for naive participants. This might indicate that their performance could have improved further with a longer familiarization.

Data collection with feedback improved the perceived controllability of the prosthesis by about 10 % on average (Figure 5). However, this result was not statistically significant, probably due to the high variability in the results. Naive participants provided extremely scattered opinions regarding the system's controllability; moreover, the average controllability reported by naives and experienced differed by about 20 %, which also contributed to the spread of the overall results. More focused questions would have been beneficial to reduce this variance. Nonetheless, the improvement reported by experienced participants exhibited a clear trend in favor of the acquisition strategies with feedback. While the average controllability reported after HLSS-DA was similar to HL-DA and higher than OL-DA, the myocontrol performance provided by HLSS-DA lay between the other two strategies and showed high variability both for naives and experienced participants. This seems to confirm, again, that the sample selection criterion has been beneficial for some of the participants and disruptive for others, and that its effectiveness requires an adequate tuning of the parameters.

Notice that the tasks perceived as most difficult were dialing a phone number and completing the pegboard game, Figure 4, which required touching small target areas with a stable pointing index finger and placing small objects in positions that were difficult to reach. Interestingly, the only task in which the improvement between OL-DA and the other strategies was statistically significant was exactly dialing the phone number, Figure 2D. Assuming the tasks' difficulty to be related to producing a firm hand gesture in challenging limb positions , then the use of feedback during data acquisition seems to improve the myocontrol robustness precisely where it is needed.

Interactivity before and during prosthetic usage lets the user develop more trust in the prosthesis through the usage of a friendly interface, dexterous but straightforward at the same time. In [19] a partially satisfactory result appear, mainly due, we speculate, to a suboptimally designed interaction protocol. All in all, however, the benefits of interaction and of HL strategies in particular are not guaranteed to transfer to disabled users and this issue must be furher investigated. We have dealt with this issue already in Gigli et al. [14], to which we refer the interested reader. The proposed acoustic feedback was designed to identify when the model prediction is wrong based on the prediction error, thus assuming that the user stably maintains the ground truth during the data acquisition. This can be guaranteed for able-bodied users thanks to proprioceptive and visual feedback of their hand configuration, but less so for amputees. A transradial amputee, particularly a naive or distracted one (in distracting conditions), might struggle to maintain a consistent muscular activation during the data acquisition, which could trigger the acoustic feedback for arm configurations where more data is not required. That circumstance is potentially disruptive for the acquisition algorithm since it may lead the user to acquire more training samples while providing a wrong muscular pattern or in arm configurations that do not need to be reinforced with more data. Users with higher amputation levels could struggle even more since they might also have problems following the proper arm trajectory.

This problem might be mitigated by inducing the user to perform more consistent and repeatable muscular activations. One could do so by exploiting bilateral mirrored training or using the prosthetic hand as a proxy for the missing limb during the data acquisition [35]. If this sort of training proves ineffective, then the feedback should be redesigned based on other metrics. Competences ranging from psychology to human-machine interfaces design, as well as a large number of focus groups and user studies, will be required to solve this problem.

4.1. Conclusion

This work shows the advantages of using interactive dynamic data acquisition strategies to train robust myocontrol models. Data acquisition for myocontrol models is traditionally performed in a non-interactive, open-loop fashion. We have proposed an interactive incremental data acquisition and model building scheme, where the myocontrol model is built incrementally in realtime during the data acquisition, while the participant receives auditory feedback about the usefulness of the currently acquired data sample. In our experiment, data acquisition strategies that guided the participant to identify and collect more training samples in problematic areas of the input space yielded better performance, both objective and subjective, and granted the participant a better understanding of the system.

Acknowledgments

This work was partially supported by the DFG project Deep-Hand, CA1389/1-2.

References

[1] Micera S, Carpaneto J and Raspopovic S 2010 "Control of hand prostheses using peripheral information" *IEEE reviews in biomedical engineering* **3** 48–68

[2] Peerdeman B, Boere D, Witteveen H, Hermens H, Stramigioli S, Rietman H, Veltink P, Misra S *et al.* 2011 "Myoelectric forearm prostheses: state of the art from a user-centered perspective." *Journal of Rehabilitation Research & Development* **48**

[3] Carey S L, Lura D J and Highsmith M J 2015 "Differences in myoelectric and body-powered upper-limb prostheses: Systematic literature review." *Journal of Rehabilitation Research & Development* **52**

[4] Fougner A, Stavdahl Ø, Kyberd P J, Losier Y G and Parker P A 2012 "Control of upper limb prostheses: Terminology and proportional myoelectric control—a review" *IEEE Transactions on neural systems and rehabilitation engineering* **20** 663–677

[5] Sensinger J W, Lock B A and Kuiken T A 2009 "Adaptive pattern recognition of myoelectric signals: exploration of conceptual framework and practical algorithms" *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **17** 270–278

[6] Fougner A, Scheme E, Chan A D, Englehart K and Stavdahl Ø 2011 "Resolving the limb position effect in myoelectric pattern recognition" *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **19** 644–651

[7] Campbell E, Phinyomark A and Scheme E 2020 "Current trends and confounding factors in myoelectric control: Limb position and contraction intensity" *Sensors* **20** 1613

[8] Woodward R B and Hargrove L J 2019 "Adapting myoelectric control in real-time using a virtual environment" *Journal of neuroengineering and rehabilitation* **16** 11

[9] Yang D, Gu Y, Jiang L, Osborn L and Liu H 2017 "Dynamic training protocol improves the robustness of pr-based myoelectric control" *Biomedical Signal Processing and Control* **31** 249–256

[10] Khushaba R N, Takruri M, Miro J V and Kodagoda S 2014 "Towards limb position invariant myoelectric pattern recognition using time-dependent spectral features" *Neural Networks* **55** 42–58

[11] Betthauser J L, Hunt C L, Osborn L E, Masters M R, Lévay G, Kaliki R R and Thakor N V 2018 "Limb position tolerant pattern recognition for myoelectric prosthesis control with adaptive sparse representations from extreme learning" *IEEE Transactions on Biomedical Engineering* **65** 770–778

[12] Scheme E, Biron K and Englehart K 2011 "Improving myoelectric pattern recognition positional robustness using advanced training protocols" *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE) pp 4828–4831

[13] Radmand A, Scheme E and Englehart K 2014 "A characterization of the effect of limb position on emg features to guide the development of effective prosthetic control schemes" *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE) pp 662–667

[14] Gigli A, Gijsberts A and Castellini C 2020 "The merits of dynamic data acquisition for realistic myocontrol" *Frontiers in Bioengineering and Biotechnology*

[15] Igual C, Pardo L A, Hahne J M and Igual J 2019 "Myoelectric control for upper limb prostheses" *Electronics* **8** 1244

[16] Antfolk C, D'Alonzo M, Rosén B, Lundborg G, Sebelius F and Cipriani C 2013 "Sensory feedback in upper limb prosthetics" *Expert review of medical devices* **10** 45–54

[17] Meattini R, Biagiotti L, Palli G and Melchiorri C 2019 "Grasp-oriented myoelectric interfaces for robotic hands: A minimal-training synergy-based framework for intent detection, control and perception" *International Workshop on Human-Friendly Robotics* (Springer) pp 110–124

[18] Hahne J M, Markovic M and Farina D 2017 "User adaptation in myoelectric man-machine interfaces" *Scientific Reports* **7** 4437

[19] Nowak M, Castellini C and Massironi C 2018 "Applying Radical Constructivism to machine learning: a pilot study in assistive robotics" *Constructivist Foundations* **13** 250–262 URL `http://constructivist.info/13/2/250.nowak`

[20] Meattini R, Nowak M, Melchiorri C and Castellini C 2019 "Automated instability detection for interactive myocontrol of prosthetic hands" *Frontiers in Neurorobotics* **13** URL `https://www.frontiersin.org/articles/10.3389/fnbot.2019.00068`

[21] Dosen S, Markovic M, Somer K, Graimann B and Farina D 2015 "Emg biofeedback for online predictive control of grasping force in a myoelectric prosthesis" *Journal of neuroengineering and rehabilitation* **12** 55

[22] Hahne J M, Dähne S, Hwang H J, Müller K R and Parra L C 2015 "Concurrent adaptation of human and machine improves simultaneous and proportional myoelectric control" *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **23** 618–627

[23] Stephens-Fripp B, Alici G and Mutlu R 2018 "A review of non-invasive sensory feedback methods for transradial prosthetic hands" *IEEE Access* **6** 6878–6899

[24] Shehata A W, Scheme E J and Sensinger J W 2018 "Improving internal model strength and performance using augmented feedback" *bioRxiv* 259754

[25] Marasco P D, Kim K, Colgate J E, Peshkin M A and Kuiken T A 2011 "Robotic touch shifts perception of embodiment to a prosthesis in targeted reinnervation amputees" *Brain* **134** 747–758

[26] Sierra González D and Castellini C 2013 "A realistic implementation of ultrasound imaging as a human-machine interface for upper-limb amputees" *Frontiers in Neurorobotics* **7** URL `https://www.frontiersin.org/articles/10.3389/fnbot.2013.00017`

[27] Gijsberts A, Bohra R, Sierra González D, Werner A, Nowak M, Caputo B, Roa M A and Castellini C 2014 "Stable myoelectric control of a hand prosthesis using non-linear incremental learning" *Frontiers in Neurorobotics* **8** URL `https://www.frontiersin.org/articles/10.3389/fnbot.2014.00008`

[28] Patel G K, Nowak M and Castellini C 2017 "Exploiting knowledge composition to improve real-life hand prosthetic control" *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25** 967–975

[29] Strazzulla I, Nowak M, Controzzi M, Cipriani C and Castellini C 2017 "Online bimanual manipulation using surface electromyography and incremental learning" *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25** 227–234 URL `https://ieeexplore.ieee.org/document/7460959`

[30] Wang S, Hsu J, Trent L, Ryan T, Kearns N, Civillico E and Kontson K 2018 "Evaluation of performance-based outcome measures for the upper limb: a comprehensive narrative review" *PM&R*

[31] Hermansson L M, Fisher A G, Bernspång B and Eliasson A C 2005 "Assessment of capacity for myoelectric control: a new Rasch-built measure of prosthetic hand control" *Journal of rehabilitation medicine* **37** 166–71

[32] Kyberd P J, Murgia A, Gasson M, Tjerks T, Metcalf C, Chappell P H, Warwick K, Lawson S E and Barnhill T 2009 "Case studies to demonstrate the range of applications of the Southampton Hand Assessment Procedure" *British Journal of Occupational Therapy* **72** 212–218

[33] Vujaklija I, Roche A D, Hasenoehrl T, Sturma A, Amsuess S, Farina D and Aszmann O C 2017 "Translating research on myoelectric control into clinics—are the performance assessment methods adequate?" *Frontiers in neurorobotics* **11** 7

[34] Chen S Y, Feng Z and Yi X 2017 "A general introduction to adjustment for multiple comparisons" *Journal of thoracic disease* **9** 1725

[35] Igual C, Pardo L A, Hahne J M and Igual J 2019 "Myoelectric control for upper limb prostheses" *MDPI Electronics* **8** 1244