Multi-grip classification-based prosthesis control with two sensors

Agamemnon Krasoulis[1,2,*], Sethu Vijayakumar[1], and Kianoush Nazarpour[2,3,*]

[1]School of Informatics, University of Edinburgh, EH8 9QD, United Kingdom
[2]School of Engineering, Newcastle University, NE1 7RU, United Kingdom
[3]Institute of Neuroscience, Newcastle University, NE2 4HH, United Kingdom
[*]Agamemnon.Krasoulis@newcastle.ac.uk; Kianoush.Nazarpour@newcastle.ac.uk

ABSTRACT

In the field of upper-limb myoelectric prosthesis control, the use of statistical and machine learning methods has been long proposed as a means of enabling intuitive grip selection and activation; yet, clinical adoption remains rather limited. One of the main causes hindering clinical translation of machine learning-based prosthesis control is the requirement for a large number of electromyography (EMG) sensors. Here, we propose an end-to-end strategy for multi-grip, classification-based prosthesis control using only two sensors, comprising EMG electrodes and inertial measurement units (IMUs). We emphasise the importance of accurately estimating posterior class probabilities and rejecting predictions made with low confidence, so as to minimise the rate of unintended prosthesis activations. To that end, we propose a confidence-based error rejection strategy using grip-specific thresholds. We evaluate the efficacy of the proposed system with real-time pick and place experiments using a commercial multi-articulated prosthetic hand and involving 12 able-bodied and two transradial (i.e., below-elbow) amputee participants. Results promise the potential for deploying intuitive, classification-based multi-grip control in existing upper-limb prosthetic systems.

Introduction

Upper-limb myoelectric prostheses are electromechanical devices that aim to partially restore the functionality and appearance of a missing limb. They typically comprise a muscular activity recording unit based on surface electromyography (EMG), an active end-effector, such as a prosthetic hand with motorised digits and a wrist rotation unit, and a processing unit that translates the recorded muscular activity information into motor commands for the end-effector. Nowadays, there exists a plethora of prosthetic devices with remarkable mechanical properties, nearly approximating the dexterity of the human hand. Nevertheless, the full mechanical capabilities of state-of-the-art prosthetic hands are seldom utilised; the ineffectiveness of the deployed control algorithms, the non-stationary nature of EMG signals, and lack of an ample number of independent muscle control sites hinder the dexterous control of myoelectric prostheses[1].

State-of-the-art active prostheses, such as the Össur i-Limb ultra[1] and Ottobock bebionic[2] hands, are typically shipped with a set of pre-programmed grip patterns. These can be utilised by the user to hold objects or perform other activities of daily living. A pair of surface EMG electrodes is commonly used to monitor the activity of flexor and extensor muscle groups, thus providing the user with control over a single degree of freedom (DOF), such as hand opening/closing within a specified grip configuration or wrist rotation. To switch between different grip patterns or functions, the user has to either perform a series of muscle co-contractions to shift through the available modules[2] or send a trigger signal associated with the desired grip, such as a double or triple impulse. This control scheme is robust but lacks intuitiveness, which in turn can lead to increased cognitive load.

Since the 1970's, significant efforts have been put into utilising computational tools from the fields of statistics, pattern recognition and machine learning to enhance the control of upper-limb myoelectric prostheses. One prominent example of this approach is the deployment of classification algorithms with the aim of increasing the intuitiveness of the control interface. The fundamental working principle of this paradigm is that features extracted from multiple EMG electrodes form motion-specific clusters in a high-dimensional space that can be used to discriminate classes of movement. Thus, to access a specific function, a user only needs to activate their muscles in a naturalistic fashion, much like they would do with an intact limb. This approach has demonstrated proof-of-principle for decoding arm movements, such as elbow and wrist flexion/extension and hand

[1]https://www.ossur.com/prosthetic-solutions/products/touch-solutions/i-limb-ultra
[2]https://www.ottobock.co.uk/prosthetics/upper_limb_prosthetics/product-systems/bebionic-hand/

opening/closing[3–5], and has recently found its way to commercial adoption[3,4]. Several studies have also used this paradigm to decode grips and gestures for intuitive prosthetic hand control. In their majority, however, they have been either limited to offline analyses[6–8] or only included able-bodied participants[9–11], with few exceptions demonstrating real-time control with amputees[12,13].

One caveat of classification-based myoelectric control is the requirement for a relatively large number of sensors[14,15] or high-density electrode arrays[16,17]. This requirement increases the overall complexity and cost of the system and reduces its practicality, due to increased weight and the burden associated with a constant need for positioning a large number of electrodes. Minimising the number of electrodes used in a myoelectric control system has been identified as one of the main challenges in the field[1]. Hence, a significant body of work has previously investigated means of achieving this goal. Exhaustive search or sequential selection algorithms have been used to identify a suitable subset from a larger pool of sensors, typically in the range 4-12[12,18–28]. Despite previous efforts, the feasibility of using a single pair of sensors, which is typically available in commercial prostheses, to control a multi-grip prosthetic hand has yet to be demonstrated.

Drastically reducing the amount of sensors used for myoelectric control may lead to a decrease in classification performance. Additionally, it has been reported that unintended prosthesis motions can lead to user frustration[29], which in turn may increase the risk of prosthesis rejection. To ensure user satisfaction, it is imperative to design fault-tolerant myoelectric controllers with the ability to reject classification predictions unless they are made with high confidence. Several post-processing strategies have been proposed for reducing the frequency of unintended prosthesis activations, including but not limited to: majority voting, whereby the control action at a given time step is affected by previous and potentially future predictions (i.e. non-causal filter)[30]; training multiple binary one-vs.-rest or one-vs.-one classifiers and rejecting classification predictions unless unanimous agreement is reached among the pool of classifiers[29,31]; threshold-based confidence rejection, whereby a classification prediction is rejected unless the respective posterior probability exceeds a fixed[4,12,32] or class-specific[26] threshold; and using an auxiliary binary classifier to estimate whether the predictions of the base motion classifier are accurate[33]. All these methods reduce the rate of unintended prosthesis activations. However, this happens at the expense of an increase in computational complexity[29,31,33], response delay[30] or a decrease in overall classification accuracy[29].

In this work, we propose a full pipeline for real-time prosthetic hand control by using a single pair of sensors comprising EMG electrodes and inertial measurement units (IMUs). Special attention is given to optimising system parameters such that the amount of unintended performed motions is minimised. To that end, a novel algorithm is introduced for selecting class-specific confidence thresholds, which is based on false positive rate (FPR) minimisation. The efficacy of the proposed system is evaluated with real-time, object pick and place experiments using a commercial prosthetic hand and involving both able-bodied and transradial amputee participants. This is the first time that machine learning-based, multi-grip prosthesis control is demonstrated in real-time with amputee subjects by using only two sensors, thus showcasing the potential for deploying the proposed methodology in existing commercial prostheses.

Results

The experimental protocol is shown in Fig. 1. Twelve able-bodied and two transradial amputee subjects took part in the experiments. Initial training data were recorded with 16 EMG-IMU sensors for able-bodied participants. For the two amputees, 12 and 13 sensors were used, respectively. For each individual, two datasets were collected in the first part of the experiments. An optimal subset of two sensors was then selected using the sequential forward selection algorithm. The selected subset was used in the second part of the experiments, where participants controlled a commercial prosthesis in real-time to lift, relocate, and release a series of objects, and finally press the space key on a computer keyboard (Fig. 2). The prosthesis was attached to a custom-built splint in the case of able-bodied subjects. The two amputee participants wore their own, but modified, sockets. More details about participant recruitment, the experimental protocol, aparatus, signal processing, prosthesis control, and the used performance metrics are provided in the Methods section.

Offline analysis

We evaluated decoding performance for the family of discriminant analysis classifiers, namely, linear discriminant analysis (LDA), regularised discriminant analysis (RDA), and quadratic discriminant analysis (QDA), with a varying number of sensors. To that end, 10-fold cross-validation was performed on the second dataset for each participant, by using a 90%-10% split. That is, the whole of the first dataset for each individual was used to train models (*training set*), 9 out of 10 repetitions of each motion from the second dataset were used as a *validation set* for sensor selection and RDA hyper-parameter optimisation, and the set of left-out repetitions from the second dataset were used as a *test set*.

The results of the classifier comparison analysis are presented in Fig. 3a. Performance was assessed using the cross-entropy loss (see Methods and equation (2)), with lower values indicating better decoding performance. In general, performance

[3] https://www.coaptengineering.com/
[4] https://www.ottobock.com/en/press/press-releases/news-detail-page_24962.html

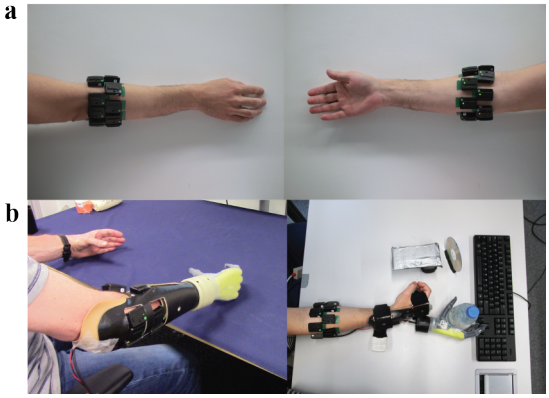

Figure 1. Sensor placement and experimental setup. **a**, Two rows of eight equally-spaced sensors were placed around the forearm of the able-bodied participants. For the two amputee subjects, 12 and 13 sensors were used, respectively. For each participant, the optimal subset of two sensors was identified and used for real-time decoding. **b**, Pictures of one amputee *(left)* and one able-bodied *(right)* participant during the training and real-time control phases of the experiment, respectively. The shown amputee participant performed bilateral mirrored movements during training. For able-bodied participants, a custom-built splint was used to support the prosthesis. The two amputee participants wore modified own sockets.

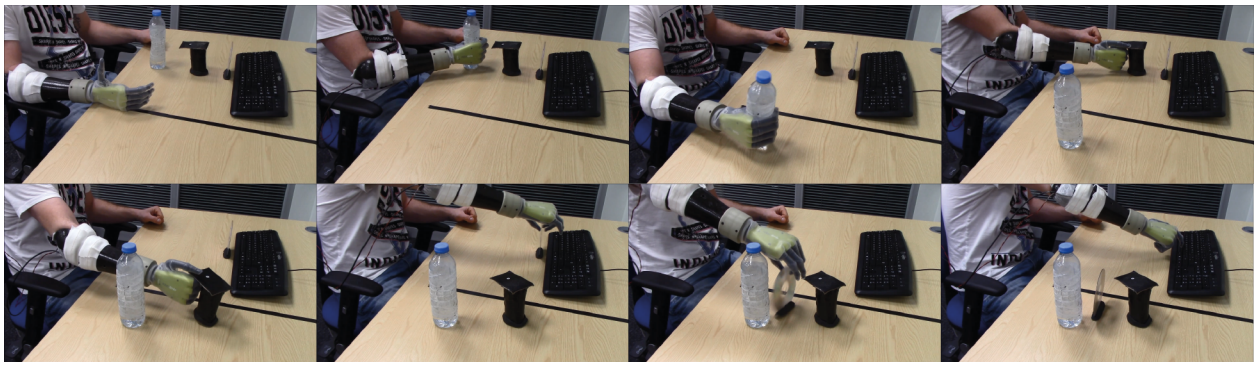

Figure 2. Real-time multi-grip prosthesis control task. A full trial is shown as a sequence of images. Participants were instructed to use the prosthetic hand to lift, relocate, and release a series of objects, and finally press the space bar on a computer keyboard. An intermediate hand "open" grip was required to switch between different grasps. The object presentation order in the shown trial was: "water bottle", "credit card simulator", and "CD"; therefore, the required sequence of prosthesis grips was: "power", "open", "lateral", "open", "tripod", "open", and "pointer". The completion time of the shown trial was 24 s. The video recording of this trial is provided as supplementary material.

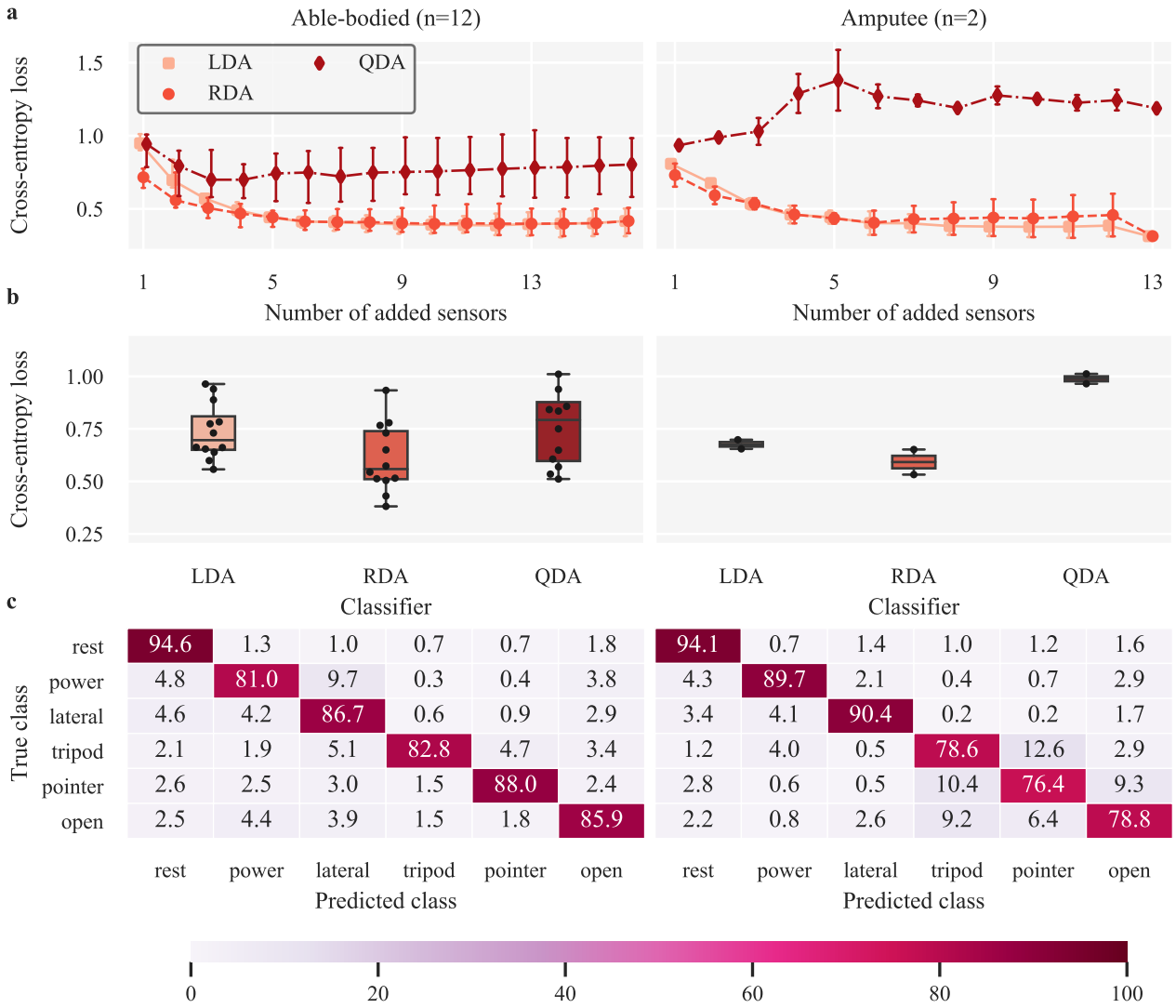

Figure 3. Offline results. Analysis was performed using 10-fold cross-validation on training data collected during the first part of the experiments. **a**, Classifier comparison for a varying number of sensors. The three considered classifiers were LDA, QDA, and RDA. The metric used for comparison was the cross-entropy loss (i.e. log-loss); lower values indicate better performance. For each configuration, the optimal subset of sensors was identified using LDA and a sequential forward selection algorithm. A total of 16 features were extracted from each sensor (seven EMG and nine inertial features). For amputees, 12 sensors were used for the second participant, thus a single data point is shown in all traces corresponding to 13 sensors and confidence intervals are not estimated in this case. Points, medians; error bars, 95% confidence intervals estimated via bootstrapping (1000 iterations). **b**, Detailed classifier comparison when using the optimal subset of two EMG-IMU sensors. Straight lines, medians; solid boxes, interquartile ranges; whiskers, overall ranges of non-outlier data; dots, individual data points. **c**, Average confusion matrices for the two participant groups using the RDA classifier and the optimal pair of sensors. Annotated scores and colour intensities indicate average per-class normalised accuracy scores.

improved as new sensors were added to the decoders and reached a plateau after the inclusion of 6-8 sensors. The RDA classifier outperformed LDA for small numbers of sensors, but the two algorithms yielded comparable scores for more than five sensors. The performance of QDA was remarkably worse than that of LDA and RDA. Interestingly, the performance of QDA deteriorated when a large number of sensors was used. The results were consistent across the able-bodied and amputee populations. Note that 12 sensors were only used for the second participant, therefore the traces corresponding to amputee subjects have a single data point for 13 sensors and confidence intervals are not estimated in this case.

The results from using the optimal subset of two EMG-IMU sensors are presented in more detail in Fig. 3b, separately for the able-bodied and amputee groups. A Friedman test revealed a statistical effect of decoding algorithm on classification performance ($p = 10^{-3}$, $n = 14$). Post-hoc pairwise comparisons showed that the RDA classifier significantly outperformed LDA and QDA ($p = 10^{-2}$ in both cases; $n = 14$), whereas LDA performed marginally, but not significantly, better than QDA ($p = 0.18$).

Average confusion matrices with the two optimally selected EMG-IMU sensors and the RDA classifier are shown for each participant group in Fig. 3c. The annotated scores indicate average per-class normalised accuracy scores for each group.

Real-time myoelectric prosthesis control experiment

Participants were instructed to modulate their muscular activity to control the prosthetic hand to pick and place three objects and press the space bar on a computer keyboard. A full trial as performed by one of the amputee subjects is shown in Fig. 2. Participants were instructed to complete the trials as fast as they could and trial timings were recorded by the experimenter. The RDA classifier was used for decoding and, for each individual, the optimal subset consisting of the two sensors identified in the precedent offline analysis was used. A grip predicted by the decoder was executed by the prosthetic hand only when the corresponding posterior probability exceeded the respective class-specific threshold. Participants were blind about the utilised decoding algorithm and the number of sensors. More detailed information about signal pre- and post-processing, feature extraction, decoding, and hyper-parameter tuning are provided in the Methods section.

The working principle of the proposed real-time prosthetic control paradigm is illustrated in Fig. 4 using a representative trial with one participant. The raw EMG signals from the two sensors selected for this subject are shown in Fig. 4a. The time series of classification predictions and prosthesis activations are shown in Fig. 4b and the temporal evolution of the class posterior probability distribution is shown in Fig. 4c. For this trial, the sequence of objects to be relocated was “card”, “bottle”, and “CD”; thus, the required sequence of hand grips was “lateral”, “open’, “power”, “open”, “tripod”, “open”, and “pointer”. It can be observed from Fig. 4b that although there was a relatively large number of incorrectly classified instances (black trace), the confidence-based rejection strategy discarded most of them as the corresponding posterior probabilities did not exceed the respective class-specific confidence thresholds (Fig. 4c). Overall, there were two unintended hand motions in this trial, marked with red ellipses in Fig. 4b, and the trial was successful with a completion time of 26.9 s.

Overall results for the 12 able-bodied and two amputee participants are reported in Fig. 5. Each subject performed 10 trials and completion rates and times were used to assess prosthetic control performance. Individual scores for the two metrics are presented on the left-hand side of Fig. 5a and 5b, respectively. Summary scores for the two participant groups are shown on the right-hand side of the graphs. The median completion rates were 95% and 85% for the able-bodied and amputee groups, respectively. Median completion times for successful trials were 37.43 and 44.28 s, respectively. Able-bodied subjects performed on average better than amputees with respect to both metrics, but a formal statistical comparison between the two groups was not possible, due to the small sample size of amputee participants ($n = 2$).

In addition to overall performance, we investigated the effect of task practice on prosthesis control and present the results of this analysis in Fig. 5c. Average completion times are plotted against trial numbers, separately for the two participant groups. We found that median completion times significantly decreased from the first two (*early*) to the last two (*late*) trials (median difference 7.78 s, $p = 0.007$; $n = 13$; Wilcoxon signed-rank test; one able-bodied participant was excluded from this analysis as they did not successfully complete any early trials). The average time elapsed between the initiation of the early and late stages of the experiments (i.e., start of the first and ninth trials, respectively) was 16.08 ± 1.39 min (mean $\pm$ standard error).

To further investigate the effect of task practice on muscular activity, we examined the temporal changes in muscle force over the course of the experiments. We computed raw EMG power for each electrode and used this measure as an approximation of muscle force. We carried out this analysis separately for the subset of EMG electrodes used for prosthesis control and the ones that were not used. We hypothesised that although participants were not explicitly told that only a subset of sensors was used for decoding, they would be able to develop an implicit model to identify the active muscle control sites. Thus, they could reduce the force exerted by those muscles that did not contribute to prosthesis control, so as to minimise overall metabolic cost whilst maintaining, or even improving, prosthesis control performance. The results of this analysis are shown in Fig. 6. Indeed, we observed a significant decrease ($p = 0.035$) in the EMG power of the subset of electrodes that did not contribute to prosthesis control, whereas no significant differences ($p = 0.14$) were observed for the subset of sensors used for real-time decoding ($n = 14$; two-sided paired t-tests). At the same time, performance in the experimental task improved as indicated by

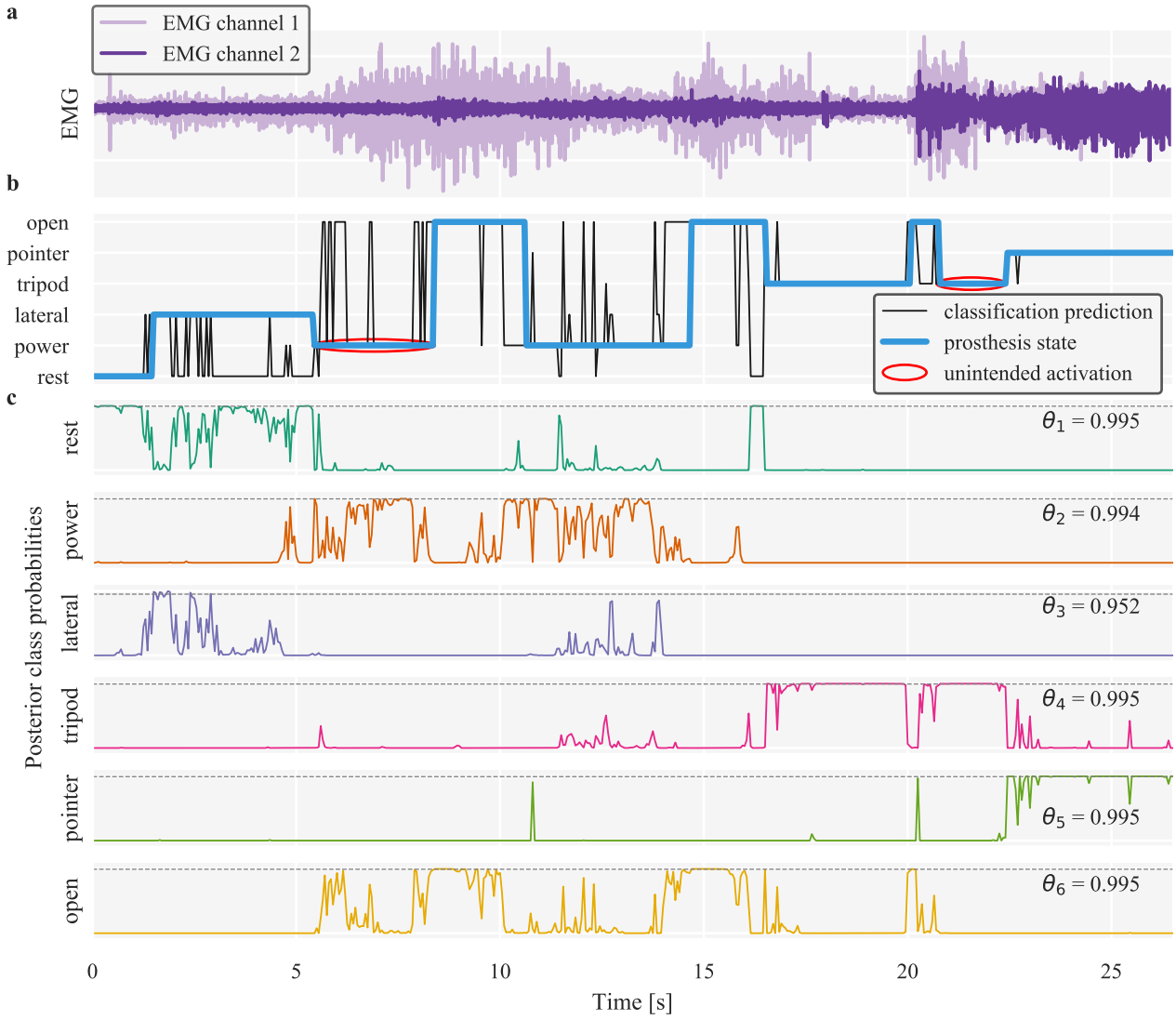

Figure 4. Working principle of the proposed real-time prosthesis control framework. A representative trial is shown for one participant. The object presentation order for this trial was "card", "bottle', and "CD"; thus, the desired grip activation sequence—including the intermediate hand "open" triggers required to switch between grasps—was "lateral", "open", "power", "open", "tripod", "open", and "pointer". **a**, Time-series of raw EMG signals from the two sensors used for the participant. **b**, Time-series of classifier predictions and temporal evolution of prosthesis state. A new classification was translated into a control action only if the corresponding posterior probability exceeded the respective class-specific threshold. When the posterior probability was lower than the corresponding rejection threshold or the "rest" class was predicted, the hand held its previous state. Unintentionally performed hand activations are marked with red ellipses. **c**, Time-series of posterior class probability distribution. Class-specific confidence thresholds are marked with dashed lines.

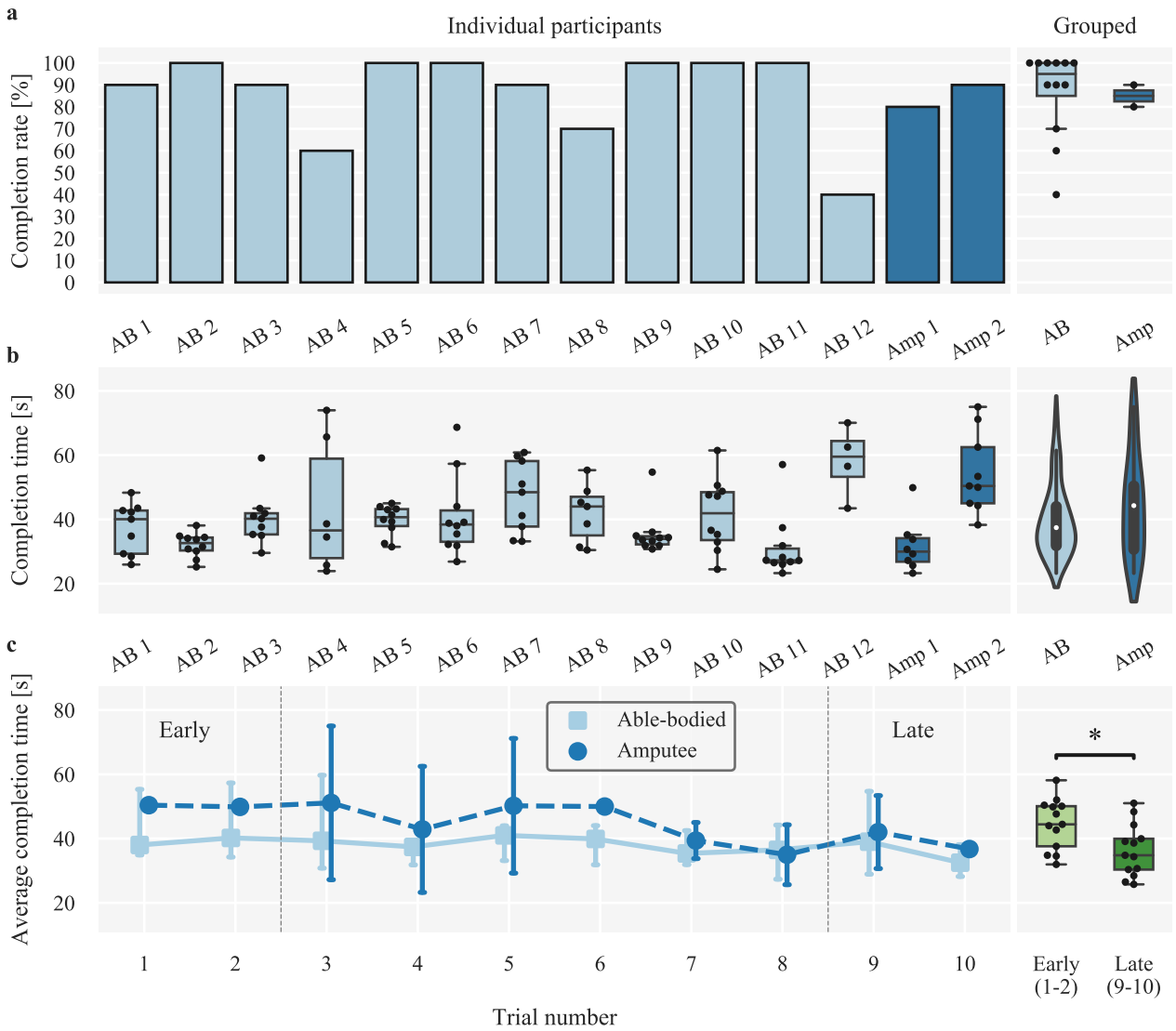

Figure 5. Real-time prosthesis control experiment results. **a**, Completion rates for individual subjects *(left)* and participant groups *(right)*. **b**, Completion times for successful trials. Violin plots, kernel density estimates of sample populations. **c** Average completion times across participant groups against trial number *(left)* and statistical comparison of early versus late trials *(right)*. Points, medians; error bars, 95% confidence intervals estimated via bootstrapping (1000 iterations); double asterisk, $p < 0.05$.

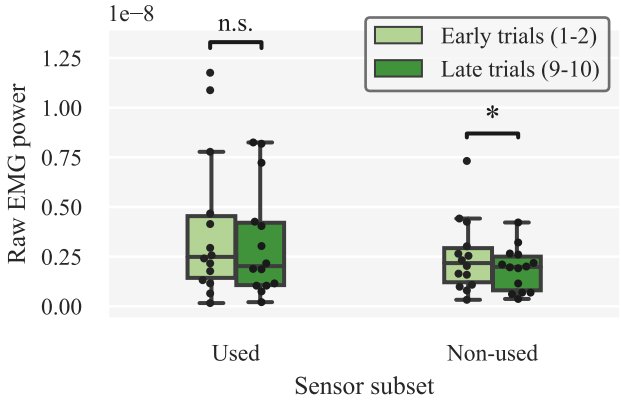

Figure 6. Effect of task practice on exerted muscle force. Statistical comparison of raw EMG power between early and late blocks for used *(left)* and non-used *(right)* sensor subsets for prosthesis control. n.s., $p > 0.05$; asterisk, $p < 0.05$.

the decrease in average completion times (Fig. 5c).

Relationship between offline and real-time performance metrics

We calculated the average offline classification accuracy and cross-entropy loss using 10-fold cross-validation for each subject. To compute balanced estimates for both measures, we removed a large fraction of the dominating "rest" class (see Methods). The balanced offline scores were then contrasted with the average completion times in the real-time control experiment in a subject-specific fashion. The results of this analysis are reported in Fig. 7 using scatter plots, where each data point corresponds to one individual. Robust linear regression fits using the Huber method ($\varepsilon = 1.345$, $n = 14$) with 95% confidence intervals are also shown in the same graphs. A positive significant ($p = 0.042$) correlation was observed between cross-entropy loss and completion time. Conversely, the correlation between classification accuracy and completion time was negative and non-significant ($p = 0.1$).

Discussion

We have demonstrated a proof-of-principle for classification-based grip control for hand prostheses using only two sensors. This finding, which to the authors' best knowledge has not been reported previously, can have a substantial impact in the field of upper-limb control, as it provides evidence that machine learning-based prosthetic control could be deployed in current commercial prosthetic hands.

Minimising the number of sensors used for myoelectric control is of significant importance for two main reasons: firstly, the use of a large number of sensors is associated with increased complexity in clinic, high power requirements and cost; secondly, a requirement for positioning a large number of electrodes can prove cumbersome for the user.

To tackle the challenging problem of controlling a prosthesis with only two sensors, we have proposed a series of novel techniques in terms of sensing, movement intent decoding, and decision making. In the sensing (i.e., control input) domain, we have made use of our previous finding that the combination of EMG and inertial measurements can improve the performance of classification-based prosthesis control[12]. Although IMUs are typically not available in commercial systems, they are relatively inexpensive components and their integration into existing systems should pose no significant challenges. In fact, some commercial devices already comprise IMUs to monitor the orientation of the prosthesis, but these are embedded in the prosthesis rather than placed on the forearm, as was the case in our study.

For movement intent decoding, we exploited the superior performance of the RDA classifier to LDA[34], especially for a small number of sensors (Fig. 3a, b). This is due to RDA producing quadratic decision boundaries in feature space, while using regularisation to avoid the overfitting issues that QDA suffers from. This was verified in the offline analysis, where we observed a decrease in QDA performance as more sensors were added to the decoding pool; this is a clear sign of overfitting, due to estimating an increasingly larger number of covariance matrix parameters without regularisation. By providing a continuum between LDA and QDA, RDA is guaranteed to achieve at least as high performance as that of the two other methods[35].

Notwithstanding the improvement provided by the use of RDA, reducing the number of sensors to only two may inevitably lead to some decrease in classification performance. To pre-empt this potential reduction, we introduced a novel confidence-based rejection strategy. This approach was based on FPR minimisation and enabled discarding predictions that were not made with high confidence. We consider this component as paramount for the efficacy of the control scheme proposed here;

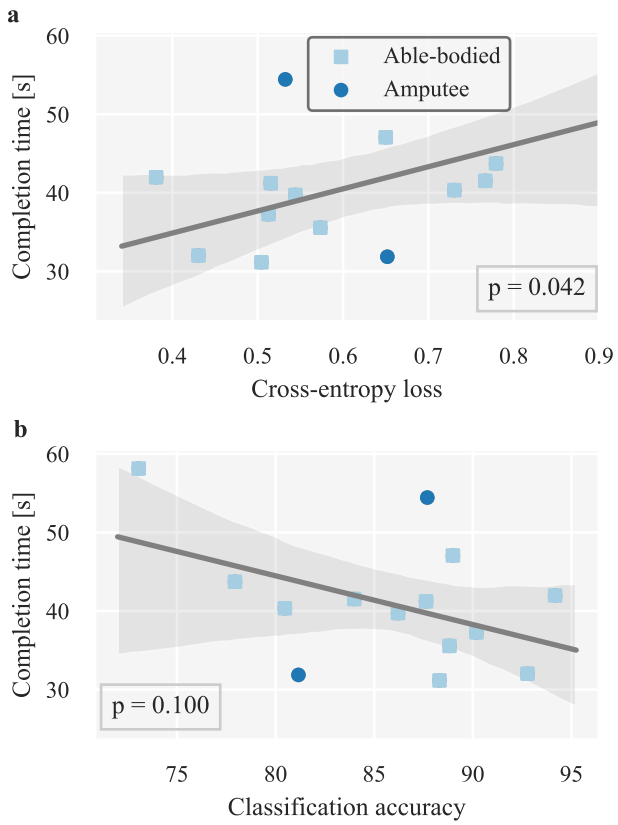

Figure 7. Relationship between offline and real-time control performance metrics. Linear relationships between average completion times and two offline metrics, i.e., **a**, cross-entropy loss and **b**, classification accuracy. Points, individual observations (i.e., participants); lines, robust linear regression fits (Huber method, $\varepsilon = 1.345$); translucent bands, 95% confidence intervals estimated via bootstrapping (1000 iterations).

without it, a substantial number of incorrect classifications might be executed by the prosthetic hand leading to performance deterioration, user frustration, and potentially damage or injury during daily life use[4,29].

Importantly, the confidence-based rejection step heavily relies on the ability to accurately estimate class posterior probabilities rather than only predicting the most probable class. Taking into consideration that many parameters need to be optimised during training, such as sensor location, classifier hyper-parameters, and rejection thresholds, a metric quantifying performance with respect to the quality of posterior probability estimates is deemed necessary. To that end, we used the cross-entropy loss, in other words, the Kullback-Leibler divergence between the true and estimated multinomial posterior class distributions (see Methods). It is worth noting that the choice of metric heavily informs the selection of hyper-parameters; different metrics can often yield utterly different results. By tuning hyper-parameters with respect to minimising cross-entropy loss, we ensured that the chosen configuration would be optimal with respect to the quality of posterior probability estimates, which plays in important role in ultimate prosthesis control given the final confidence-based rejection stage. Unfortunately, this aspect is often not considered in the field of pattern recognition-based myoelectric control; while most studies involving real-time experiments include some kind of confidence-based rejection, the ability of the decoding algorithm to produce accurate posterior probability estimates is, with very few exceptions[36], rarely assessed. Instead, the large majority of studies is predominantly concerned with maximising classification accuracy, despite increasing evidence showing that this metric is not well correlated with real-time control performance[37–39]. In our real-time experiments, we found that average completion times were more strongly correlated with cross-entropy loss than with classification accuracy (Fig. 7). In both cases, however, the data points corresponding to the two amputee participants fell outside the respective confidence intervals. Further investigation with a larger number of amputees may be required to verify whether cross-entropy loss or other probability-related metrics, such as area under the receiver operating charachteristic (ROC) curve, can provide a reliable estimator of real-time prosthesis control performance.

Moreover, paramount to maximising decoding performance is tuning the RDA regularisation hyper-parameter in a systematic way, for example, using holdout or cross-validation. In addition to the regularisation hyper-parameter, the sensor subset and rejection thresholds can influence the final prosthesis control. Therefore it is crucial to tune them by evaluating performance on a different dataset than the one used for classifier training to avoid introducing biases[40].

For rejection threshold selection using ROC curve analysis, the most commonly used strategies involve either maximising the vertical distance from a random classifier or minimising the distance from an ideal classifier[41]. However, neither of the two methods impose a constraint on the resulting FPR. This was regarded as a high priority in this work, given the high associated cost of false positive activations which translate into unintended hand activations. To address this issue, thresholds were selected such that the true positive rate (TPR) was maximised, while constraining FPR to be kept lower than a cut-off threshold (not to be confused with the prediction rejection threshold). The value for the FPR cut-off threshold was set empirically during pilot trials. An interesting avenue for future research would be to attempt to systematically identify the optimal FPR cut-off threshold during real-time myoelectric control. One possible way to achieve this could be by giving the user control over the threshold value, for example via a knob switch, and asking them to select it according to their individual preference. It shall be interesting to investigate whether a shared preference pattern can be observed across different individuals.

Various algorithms have been previously proposed for EMG channel selection, including, but not limited to, exhaustive search (i.e., brute force)[18,20], sequential forward[19,23,25,42] or backward selection[27], and independent component analysis[22]. Here, we employed the standard sequential forward selection algorithm, mainly because of its speed and efficiency during training. It has been previously demonstrated that despite its simplicity, it can outperform more sophisticated methods, such as the Lasso algorithm[25]. An alternative would have been to optimally select the pair of sensors using exhaustive search. However, in our experience, this approach can only offer a marginal improvement in performance, if any, at the expense of a substantial increase in computational complexity; the number of search iterations scales quadratically in the number of sensors with exhaustive search, as opposed to linearly with sequential selection.

In our proposed pipeline, the LDA classifier was used to assess the predictive performance of the set of candidate sensors within each iteration. Once the sensor selection algorithm terminated, the RDA regularisation hyper-parameter was optimised for the selected subset of sensors. This approach might have produced slightly suboptimal results, since the interaction effect of sensor selection and regularisation constant was not considered. Concurrently optimising for the two parameters would have, however, incurred prohibitive training times and was thus not considered. One possible alternative would be to tune these two parameters simultaneously by using state-of-the-art function optimisation methods, such as Bayesian optimisation[43].

During offline analysis, we found that the optimal number of sensors for classification may lie in the range of five to seven (Fig. 3a). This is in agreement with previous work investigating EMG channel reduction[12,19–26,28]. Bearing in mind clinical applications, we sought to investigate whether machine learning-based prosthesis control could be feasible with only two sensors. Contrasting completion times between the current and our earlier work, where we have used an average of four to six sensors in a similar task[12], we note that performance between the two conditions was comparable. A direct comparison between these two studies is, however, not feasible due to differences in various experimental design parameters, such as number of trials, length of processing window, and confidence-based rejection strategy.

We did not observe any patterns shared across the participants in terms of the position of the selected sensors. From a clinical point of view, this finding suggests that a subject-specific approach for sensor position identification may be required. One possible solution to this problem is to adopt the approach considered here, that is, record muscular activity from many sites during an initial screening, and subsequently identify the optimal sensor positions based on a search algorithm. This procedure should however precede socket fabrication, which requires that sensor positions be already established.

It has been previously demonstrated that myoelectric control performance with intuitive, classification-based decoders can increase over time with user practice[18,44,45]. In agreement with previous reports, a significant decrease was observed in completion times between early and late trials (Fig. 5c). The testing phase of the experiment lasted on average 20 minutes, thus it is reasonable to expect that performance can potentially improve with daily use, provided that the effect of exogenous parameters such as skin condition and sensors position is controlled. Interestingly, in parallel with task performance improvement, we observed a decrease in average raw EMG power of the sensor subset that was not used for decoding and hence prosthesis control (Fig. 6). It is therefore likely that, despite having not been explicitly made aware of the low-dimensional mapping between muscular activity and prosthesis control, participants were able to develop an internal model of this mapping and adapted their control strategy to simultaneously maximise performance and minimise limb impedance[46].

In conclusion, we have provided a pre-clinical proof-of-principle for classification-based multi-grip myoelectric control using only two sensors. Control of a state-of-the-art commercial prosthesis was demonstrated in real-time with able-bodied participants and transradial amputees. The proposed paradigm has the potential to transform existing upper-limb myoelectric systems to support intuitive, machine learning-based multi-grip prosthesis control. A full clinical translation of the proposed paradigm will require an embedded implementation, which will also facilitate further investigation of the long-term system viability and performance. In the future, it shall be valuable to compare this paradigm to clinical standards, such as myoelectric mode switching and body-powered prostheses.

Methods

Participant recruitment

Twelve able-bodied (10 male, two female; 10 right-hand, two left-hand dominant; median age 28 years) and two right-hand amputee subjects were recruited. Both amputee participants were right-hand dominant prior to amputation. Some of the able-bodied and both amputee subjects had previously taken part in classification-based myoelectric control experiments[12]. All experiments were performed in accordance with the Declaration of Helsinki and were approved by the local Ethics Committees of the School of Informatics, University of Edinburgh and School of Engineering, Newcastle University. All participants read an information sheet and gave written informed consent prior to the experiments.

Signal acquisition and socket fitting

For the able-bodied group, 16 EMG-IMU Delsys® Trigno™ IM sensors were placed on the participants' forearm arranged in two rows of eight equally spaced sensors each (Fig. 1). For the two amputee participants, 13 and 12 sensors were used, respectively, due to limited space availability. The sensors were placed on the able-bodied participants' dominant arm, whereas for amputees they were placed on the subjects' stump (right arm in both cases). Prior to sensor placement, the participants' skin was cleansed using 70% isopropyl alcohol. Elastic bandage was used to secure the sensor positions throughout the experimental sessions. Following sensor placement, we verified the quality of all EMG channels by visual inspection. The hardware sampling rates for EMG and inertial data were 1111 Hz and 128 Hz, respectively. Readings from IMUs were used in their raw format, therefore no calibration was required.

A splint and custom built sockets were designed to accommodate the prosthetic hand that was used in the experiments (Fig. 1 and Fig. 2). For the able-bodied group, the same splint was used for all participants and was adjusted for individual subjects using Velcro straps. The two amputee participants wore their own sockets, which were modified to accommodate the EMG-IMU sensors. The prosthesis was fitted on the dominant arm of the able-bodied subjects. For the two amputee participants, it was fitted on their stump.

Robotic hand

The Touch Bionics robo-limb™ hand was used in the experiments, which is an externally-powered, underactuated (11 DOFs, 6 degree of activations (DOAs)), anthropomorphic hand. This model is identical to the commercially available "i-Limb ultra" model. It comprises 5 motors controlling the flexion/extension of the digits and an additional motor controlling the rotation of the thumb. The hand operates under 7.4 V nominal voltage with a maximum current consumption of 7 A. During all experiments, the hand was externally powered with a doubly-insulated power supply unit, certified for medical experiments.

The robo-limb can be controlled by a computer via a CAN bus interface in an open-loop fashion. The control commands take the form "`ID - Action - PWM`": `ID` specifies the desired DOA to be activated (0-6); `Action` indicates the desired motion

(open/close/stop); and `PWM` specifies the desired pulse width modulation level to be applied to the motor, in the range $[10, 127]$, which controls the digit movement velocity.

Grip control was implemented for the hand using a set of pre-defined digit activation sequences. A grip was executed only if the most recent grip command had finished execution. In any other case, that is, if a new command was issued while the most recent one was still being executed, the newly issued command would be discarded.

Behavioural task

The participants sat comfortably on an office chair. Each session comprised a calibration and a real-time control phase. In the calibration phase, subjects were instructed to perform five motions presented to them on a computer monitor: power grasp, lateral grasp, tripod grasp, index pointer, and hand opening. In the real-time control phase, they were required to use the prosthetic hand to grasp, relocate (approximately 50 cm away from their initial position), and release three objects, and finally press the space bar on a computer keyboard. For each calibration trial, participants were instructed to execute the respective movement at a moderate speed. Once they had executed the desired grip, they were required to perform a dynamic spatial movement to imitate object relocation, thus covering with their (residual) arm the region of interest[12,47]. Calibration trials lasted 5 s each and were interleaved with 3 s of rest. All collected data were included unmodified in the subsequent analysis steps. That is, no segments of activity were removed from either the start or end of the trials. Two separate blocks of data were recorded, each one comprising 10 consecutive repetitions for each grip. The two datasets were subsequently used as *training* and *validation* sets, respectively. One of the amputee participants performed bilateral mirrored movements during the calibration phase (Fig. 1).

The following four objects were used in the experiments: a water bottle, a credit card simulator, a CD, and a computer keyboard. The corresponding prosthesis grips were power, lateral, tripod, and pointer (Fig. 2). An intermediate "open" trigger was required to switch between different grips. When the "rest" class was predicted by the decoders, there was no movement and the hand held its previous state. Depending on their laterality, participants were instructed to move the objects away from their point of reference; from centre to right for the right-handed able-bodied and the two amputee subjects, and from centre to left for the left-handed able-bodied subjects. Trials were considered successful if all objects were relocated and the space key was pressed within 75 s. In case of an object drop during relocation, the trial would be interrupted and considered as unsuccessful. The number of trials per subject was set to 10 and participants were given 45 s of rest in-between consecutive trials. The object presentation order was pseudo-randomised and counter-balanced across participant groups.

Performance assessment

For offline analyses, the cross-entropy loss —also known as logistic loss or log-loss— was used to evaluate decoding performance. The cross-entropy loss is closely related to the *Kullback-Leibler divergence* between the empirical and estimated distributions of a discrete random variable. Let $y \in \{1, \ldots, C\}$ denote a discrete target variable which is encoded as a "one-of-K" binary indicator matrix Y of dimensionality $N \times C$, such that:

$$y_{i,c} = \begin{cases} 1, & \text{if sample } i \text{ belongs to class } c, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The multi-class cross-entropy loss is then defined as follows:

$$\text{cross-entropy} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \log\left(\hat{p}_{i,c}\right), \tag{2}$$

where $\hat{p}_{i,c}$ denotes the posterior probability that sample i belongs to class c. In the ideal case —that is, when all examples in a dataset are correctly classified and the corresponding posterior probabilities are exactly 1— the cross-entropy loss is equal to 0. On the other hand, there is no lower bound for cross-entropy loss; poor probability predictions can yield arbitrarily low (i.e., large negative) scores. Prior to performance evaluation, the distribution of validation/test samples was balanced by removing a large fraction of the samples corresponding to the dominating "rest" class.

To evaluate prosthetic control performance in the real-time experiments, we adopted the following two commonly used task-related metrics: completion rate, that is, the ratio of successful to total number of trials; and completion time, defined as the time taken to accomplish a successful trial.

Signal pre-processing

Power line interference was suppressed from the myoelectric signals by applying a Hampel filter. To remove motion artefacts, band-pass filtering was applied in the range $[10, 500]$ Hz using a 4th-order Butterworth filter. Myoelectric and inertial signals were synchronised via upsampling to 2 kHz and linear interpolation.

By using a sliding window approach, we extracted four time-domain features from each EMG channel, namely, waveform length, Wilson amplitude, log-variance, and 4th-order auto-regressive coefficients. For inertial data, the mean value within the processing window was used. The length of the sliding window was set to 128 ms and the increment to 50 ms (60% overlap). The total number of extracted features was 256 for able-bodied subjects, 208 for the first, and 192 for the second amputee participant, respectively (i.e., 16, 13, or 12 sensors $\times$ 16 features/sensor). The columns of the design matrix (i.e., input features) were standardised via mean subtraction and inverse standard deviation scaling.

Stimulus presentation timings were recorded using high resolution timestamps. We used a post-hoc relabelling procedure based on EMG-stimulus alignment to refine the exact motion timings and target vector labels for each subject and trial[48].

Sensor selection

For each subject, two EMG-IMU sensors were selected from the full set by using a sequential forward selection algorithm. As part of this procedure, we fitted LDA models using the training set and assessed performance on the validation set using cross-entropy loss. In each iteration of the search algorithm, a sensor could be added to the selected subset using all 16 associated features.

Classifier training and hyper-parameter optimisation

For movement intent decoding, we used the RDA classifier, which offers a continuum between LDA and QDA by fitting class-specific covariance matrices that are regularised toward the pooled covariance matrix[35]. To tune the regularisation hyper-parameter, we used a line search in the range $[0,1]$ with a step size of 0.025, and selected the parameter value that yielded the lowest cross-entropy loss score on the validation set. Following hyper-parameter optimisation, the training and validation sets were merged and used to train final models. Model training and hyper-parameter optimisation were performed in a subject-specific fashion.

Confidence-based rejection and threshold selection

Classifier predictions were post-processed using confidence-based rejection. That is, predictions were discarded unless the corresponding posterior probabilities exceeded pre-defined, class-specific thresholds. The rejection thresholds were selected by using ROC analysis on the validation set. For that purpose, multiple one-vs.-all RDA classifiers were trained and the corresponding FPR and TPR scores were computed for threshold values in the range $[0,1]$. The rejection threshold for each class was selected such that the TPR was maximised, while the respective FPR was constrained to be smaller than a cut-off value, set *a priori* to $5 \cdot 10^{-4}$. This was done in order to minimise the number of false positives that would translate into unintended hand motions. To avoid setting thresholds extremely close to 1 for well-separated classes, which would then dramatically reduce the respective TPR during real-time control, we introduced an upper-bound for the thresholds and set it empirically to 0.995. In mathematical terms, the strategy for class-specific confidence threshold selection can be summarised as follows:

$$\hat{\theta}_c = \min\left\{\max_{\theta_c} TPR(\theta_c) : FPR(\theta_c) < 5 \cdot 10^{-4}, \quad 0.995\right\}, \tag{3}$$

where $\hat{\theta}_c$ denotes the confidence threshold for class c. A representative example of this procedure for one subject and a single movement class is shown in Fig. 8. In this example, the selected threshold value was 0.990 and the corresponding TPR was 0.439.

Statistical analysis

All statistical comparisons were performed using non-parametric tests. Two-sided Wilcoxon signed-rank tests were used for pairwise comparisons and multiple comparison tests were corrected using the Šidák method. The statistical significance level was set to $\alpha = 0.05$.

References

1. Farina, D. *et al.* The extraction of neural information from the surface EMG for the control of upper-limb prostheses: emerging avenues and challenges. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **22**, 797–809 (2014).
2. Parker, P. A., Englehart, K. B. & Hudgins, B. S. Myoelectric signal processing for control of powered limb prostheses. *J. Electromyogr. Kinesiol.* **16**, 541–548 (2006).
3. Englehart, K. B., Hudgins, B. S., Parker, P. A. & Stevenson, M. Classification of the myoelectric signal using time-frequency based representations. *Med. Eng. & Phys.* **21**, 431–438 (1999).

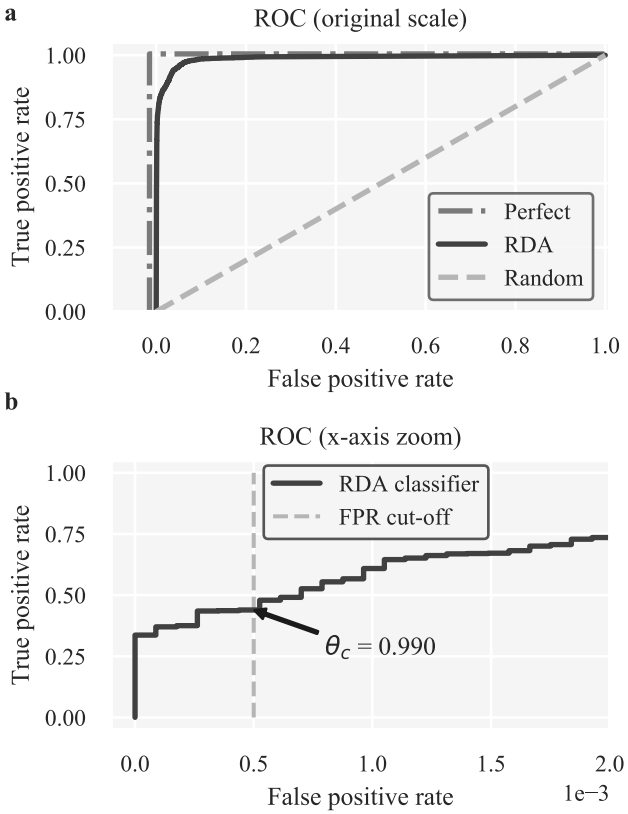

Figure 8. Confidence threshold selection. To select class-specific confidence thresholds, the multi-class classification problem needs to be transformed into multiple one-vs.-rest problems so that individual ROC curves can be computed. **a**, ROC curves shown for perfect, random, and a representative RDA classifier used in the experiments. **b** The class-specific confidence thresholds are defined by the points where the respective one-vs.-all ROC curves cross the FPR cut-off value of 5×10^{-4}.

4. Scheme, E. J., Hudgins, B. S. & Englehart, K. B. Confidence-based rejection for improved pattern recognition myoelectric control. *IEEE Transactions on Biomed. Eng.* **60**, 1563–1570 (2013).

5. Scheme, E. J. & Englehart, K. B. Validation of a selective ensemble-based classification scheme for myoelectric control using a three-dimensional fitts' law test. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **21**, 616–623 (2013).

6. Castellini, C., Gruppioni, E., Davalli, A. & Sandini, G. Fine detection of grasp force and posture by amputees via surface electromyography. *J. Physiol.* **103**, 255–262 (2009).

7. Fligge, N., Urbanek, H. & van der Smagt, P. Relation between object properties and EMG during reaching to grasp. *J. Electromyogr. Kinesiol.* **23**, 402–410 (2013).

8. Phinyomark, A. *et al.* EMG feature evaluation for improving myoelectric pattern recognition robustness. *Expert. Syst. with Appl.* **40**, 4832–4840 (2013).

9. Liu, L., Liu, P., Clancy, E. A., Scheme, E. J. & Englehart, K. B. Electromyogram whitening for improved classification accuracy in upper limb prosthesis control. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **21**, 767–774 (2013).

10. Ortiz-Catalan, M., Håkansson, B. & Brånemark, R. Real-time and simultaneous control of artificial limbs based on pattern recognition algorithms. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **22**, 756–764 (2014).

11. Gailey, A., Artemiadis, P. & Santello, M. Proof of concept of an online EMG-based decoding of hand postures and individual digit forces for prosthetic hand control. *Front. Neurol.* **8** (2017).

12. Krasoulis, A., Kyranou, I., Erden, M. S., Nazarpour, K. & Vijayakumar, S. Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements. *J. NeuroEngineering Rehabil.* **14**, 71 (2017).

13. Batzianoulis, I., Krausz, N. E., Simon, A. M., Hargrove, L. & Billard, A. Decoding the grasping intention from electromyography during reaching motions. *J. NeuroEngineering Rehabil.* **15**, 57 (2018).

14. Tenore, F. *et al.* Decoding of individuated finger movements using surface electromyography. *IEEE Transactions on Biomed. Eng.* **56**, 1427–1434 (2009).

15. Al-Timemy, A. H., Bugmann, G., Escudero, J. & Outram, N. Classification of finger movements for the dexterous hand prosthesis control with surface electromyography. *IEEE J. Biomed. Heal. Informatics* **17**, 608–618 (2013).

16. Geng, W. *et al.* Gesture recognition by instantaneous surface EMG images. *Sci. Reports* **6**, 36571 (2016).

17. Khushaba, R. N., Al-Timemy, A. H., Al-Ani, A. & Al-Jumaily, A. A framework of temporal-spatial descriptors based feature extraction for improved myoelectric pattern recognition. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **25**, 1821–1831 (2017).

18. Hargrove, L., Englehart, K. B. & Hudgins, B. S. A comparison of surface and intramuscular myoelectric signal classification. *IEEE Transactions on Biomed. Eng.* **54**, 847–853 (2007).

19. Huang, H., Zhou, P., Li, G. & Kuiken, T. A. An analysis of EMG electrode configuration for targeted muscle reinnervation based neural machine interface. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **16**, 37–45 (2008).

20. Li, G., Schultz, A. E. & Kuiken, T. A. Quantifying pattern recognition-based myoelectric control of multifunctional transradial prostheses. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **18**, 185–192 (2010).

21. Fougner, A. L., Stavdahl, Ø. & Kyberd, P. J. System training and assessment in simultaneous proportional myoelectric prosthesis control. *J. NeuroEngineering Rehabil.* **11**, 75 (2014).

22. Naik, G. R., Al-Timemy, A. H. & Nguyen, H. T. Transradial amputee gesture classification using an optimal number of sEMG sensors: an approach using ICA clustering. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **24**, 837–846 (2016).

23. Adewuyi, A. A., Hargrove, L. J. & Kuiken, T. A. An analysis of intrinsic and extrinsic hand muscle EMG for improved pattern recognition control. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **24**, 485–494 (2016).

24. Muceli, S., Jiang, N. & Farina, D. Extracting signals robust to electrode number and shift for online simultaneous and proportional myoelectric control by factorization algorithms. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **22**, 623–633 (2014).

25. Hwang, H.-J., Hahne, J. M. & Müller, K.-R. Channel selection for simultaneous and proportional myoelectric prosthesis control of multiple degrees-of-freedom. *J. Neural Eng.* **11**, 056008 (2014).

26. Menon, R. *et al.* Automatic misclassification rejection for LDA classifier using ROC curves. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 482–485 (IEEE, 2015).

27. Clancy, E. A., Martinez-Luna, C., Wartenberg, M., Dai, C. & Farrell, T. R. Two degrees of freedom quasi-static EMG-force at the wrist using a minimum number of electrodes. *J. Electromyogr. Kinesiol.* **34**, 24–36 (2017).

28. Samadani, A. Emg channel selection for improved hand gesture classification. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4297–4300 (IEEE, 2018).

29. Hargrove, L. J., Scheme, E. J., Englehart, K. B. & Hudgins, B. S. Multiple binary classifications via linear discriminant analysis for improved controllability of a powered prosthesis. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **18**, 49–57 (2010).

30. Englehart, K. B. & Hudgins, B. S. A robust, real-time control scheme for multifunction myoelectric control. *IEEE Transactions on Biomed. Eng.* **50**, 848–854 (2003).

31. Scheme, E. J., Englehart, K. B. & Hudgins, B. S. Selective classification for improved robustness of myoelectric control under nonideal conditions. *IEEE Transactions on Biomed. Eng.* **58**, 1698–1705 (2011).

32. Robertson, J., Scheme, E. & Englehart, K. Effects of Confidence-Based Rejection on Usability and Error in Pattern Recognition-Based Myoelectric Control. *IEEE J. Biomed. Heal. Informatics* **In Press** (2018).

33. Amsüss, S. *et al.* Self-correcting pattern recognition system of surface EMG signals for upper limb prosthesis control. *IEEE Transactions on Biomed. Eng.* **61**, 1167–1176 (2014).

34. Krasoulis, A., Nazarpour, K. & Vijayakumar, S. Use of regularized discriminant analysis improves myoelectric hand movement classification. In *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)*, 395–398 (2017).

35. Friedman, J. H. Regularized discriminant analysis. *J. Am. Stat. Assoc.* **84**, 165–175 (1989).

36. Scheme, E. & Englehart, K. A comparison of classification based confidence metrics for use in the design of myoelectric control systems. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 7278–7283 (IEEE, 2015).

37. Simon, A. M., Hargrove, L. J., Lock, B. A. & Kuiken, T. A. Target achievement control test: evaluating real-time myoelectric pattern-recognition control of multifunctional upper-limb prostheses. *The J. Rehabil. Res. Dev.* **48**, 619 (2011).

38. Ortiz-Catalan, M., Rouhani, F., Brånemark, R. & Håkansson, B. Offline accuracy: a potentially misleading metric in myoelectric pattern recognition for prosthetic control. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1140–1143 (2015).

39. Vujaklija, I. *et al.* Translating research on myoelectric control into clinics—Are the performance assessment methods adequate? *Front. Neurorobotics* **11** (2017).

40. Domingos, P. A few useful things to know about machine learning. *Commun. ACM* **55**, 78 (2012).

41. Gönen, M. *Analyzing receiver operating characteristic curves with SAS* (SAS Institute, 2007).

42. Geng, Y., Zhang, X., Zhang, Y.-T. & Li, G. A novel channel selection method for multiple motion classification using high-density electromyography. *Biomed. Eng. Online* **13**, 102 (2014).

43. Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian optimization of machine learning algorithms. *Adv. Neural Inf. Process. Syst.* 2951–2959 (2012).

44. Bunderson, N. E. & Kuiken, T. A. Quantification of feature space changes with experience during electromyogram pattern recognition control. *IEEE Trans. Neural Syst. Rehabil. Eng.* **20**, 239–246 (2012).

45. Powell, M. A., Kaliki, R. R. & Thakor, N. V. User training for pattern recognition-based myoelectric prostheses: improving phantom limb movement consistency and distinguishability. *IEEE Transactions on Neural Syst. Rehabil. Eng.* **22**, 522–532 (2014).

46. Burdet, E., Osu, R., Franklin, D. W., Milner, T. E. & Kawato, M. The central nervous system stabilizes unstable dynamics by learning optimal impedance. *Nature* **414**, 446 (2001).

47. Radmand, A., Scheme, E. J. & Englehart, K. B. On the suitability of integrating accelerometry data with electromyography signals for resolving the effect of changes in limb position during dynamic limb movement. *J. Prosthetics Orthot.* **26**, 185–193 (2014).

48. Kuzborskij, I., Gijsberts, A. & Caputo, B. On the challenge of classifying 52 hand movements from surface electromyography. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4931–4937 (IEEE, 2012).

Acknowledgements

The authors are grateful to the two amputee volunteers for taking part in this study. The authors also thank Sigrid Dupan for providing comments on an earlier version of the manuscript. AK was supported in part by grants EP/F500386/1 and BB/F529254/1 for the University of Edinburgh School of Informatics Doctoral Training Centre in Neuroinformatics and Computational Neuroscience from the UK Engineering and Physical Sciences Research Council (EPSRC), UK Biotechnology and Biological Sciences Research Council (BBSRC), and the UK Medical Research Council (MRC). AK is currently supported by EPSRC grant EP/R004242/1. SV is supported by the Microsoft Research RAEng. Fellowship and EPSRC grant EP/L016834/1. KN is supported by EPSRC grants EP/N023080/1 and EP/R004242/1.

Author contributions statement

AK, SV, and KN designed the experiments. AK implemented the experimental protocol. AK and KN conducted the experiments. AK analysed the data, prepared figures, and wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Correspondence and requests for materials should be addressed to A.K. and K.N.