Supplementary Material

Contents

Pre-surgical Screening Protocol

Briefly, individuals with C5 quadriplegia who called or visited The Miami Project to Cure Paralysis or who had registered with the research center with an interest to participate in research were offered the opportunity to participate in this screening study.

Each participant underwent screening measurements and testing during a 1-3-month evaluation period (depending schedule and availability). A total of up to 16 sessions were required for subjects to be considered for the surgical protocol. During each session the ability of the subject to reliably trigger electrical stimulation of the hand using motor imagery was tested. A subset of the results comparing the performance of healthy volunteers and SCI subjects has been separately published[1]. Upon completing these tests, the subject underwent a brain fMRI study to characterize the ability of motor imagery related to hand movement to lead to changes in the BOLD signal.

TABLE S1 SURGICAL PROTOCOL MAJOR INCLUSION CRITERIA

Inclusion Criteria	Measure	Rationale
Age $\geq$ 22	Years of age	Higher rates of neurological recovery in adults and better potential for rehabilitation
Age $\leq$ 50	Years of age	Lower risk of complications
AIS Grades A & B	Neurological exam	Standard neurologic assessment for spinal cord injury
Level of injury C5 motor complete	Neurological exam	The appropriate injury level for measuring detectable restoration of both triceps and hand function
Local community dwelling	Proof of local community address	Higher compliance to weekly follow-up visits

Stable chronic injury	1-15 years post injury	Suitability for efficacy measurements while excluding complications that develop with excessive time post-injury
Stable health status and upper extremities	No significant joint contractures at the elbow, wrists, or hands	Minimize interferences with the ability to perform outcome measure tests like transfer or pinch
Participation in Clinical EEG Protocol	Successful screening and assessments outlined in Clinical EEG Protocol	Selection of candidates screened in Clinical EEG Protocol ensures participants possess ability to trigger orthosis with motor imagery

TABLE S2 SURGICAL PROTOCOL MAJOR EXCLUSION CRITERIA

Exclusion Criteria	Measure	Rationale
Coagulopathy	Lab test	Higher risks of complications
Anticoagulation	Lab test	Higher risks of surgical complications
Pregnancy	Urine or serum pregnancy test	Risk to fetus

Additional Exclusion Criteria

1. Subjects with severe non-CNS injury or serious concurrent medical issues.
2. Subjects with metal prosthetics.
3. Subjects with tendon transfers.
4. Subjects with a history of cardiac arrhythmia.
5. Subjects with cognitive issues. During Screening, subjects will undergo a Neuropsychological testing and use the Mini-Mental Status has a guide to severity of dementia. A MMSE score of ≤ 15 will be an absolute exclusion. Anything above that score but still abnormal will be a relative contraindication requiring discussion amongst the research team.
6. Subjects who have made a suicide attempt or are severely depressed. Severe depression should be defined based on a depression assessment scale.
7. Subjects with peripheral nerve damage that will affect planned investigational testing will be excluded.
8. Subjects with medical condition that requires regular post-implant MRIs.
9. Subjects who suffer from claustrophobia (contraindication for fMRI).
10. Subjects who have a life expectancy of less than 2 years.

11. Subjects diagnosed with peripheral polyneuropathy.
12. Subjects with another implanted stimulator (e.g., pacemaker, defibrillator, cochlear implant, neurostimulator, etc.)
13. Subjects who have been, or are currently, enrolled in another investigational study.
14. Subjects who have skin ulcerations, a history of poor wound healing, an active infection, or significant pain in the lower extremity that is being treated.
15. Subjects unable to give informed consent.
16. Subjects who are prisoners or wards of the state.
17. Subjects who are pregnant or planning to become pregnant.
18. Subjects that speak languages without local site level expertise for translation and verbal communication.
19. Screening Study Specific Exclusion
    a. fMRI does not show reproducible hand/arm activation
    b. External triceps stimulation does not produce consistent, adequate, and reproducible contraction
    c. EEG BCI studies do not lead to viable control

Surgical Study Timeline

The presented subject was enrolled in the study on November 2, 2018 and surgical implantation occurred on November 30, 2018. A timeline of the important study events is given Figure S1.

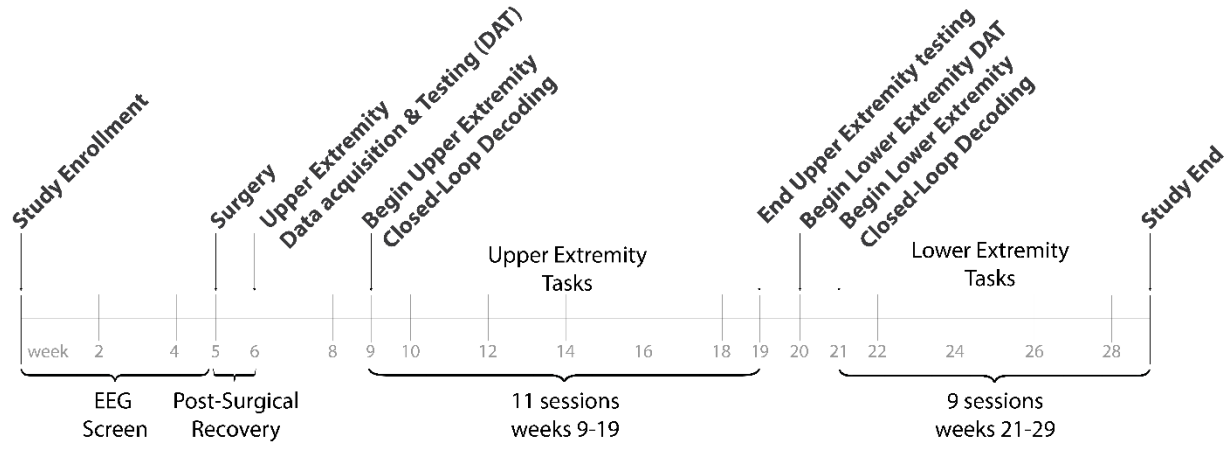


FIGURE S1 STUDY TIMELINE. DAY 0 OF THE STUDY BEGINS ON THE DATE THE SUBJECT SIGNED INFORMED CONSENT TO BEGIN THE SCREENING PROTOCOL WITH EEG.

Decoding upper extremity movement intent

After surgical recovery, the subject came to the laboratory 2-3 times per week for 1-2 hours at a time. A timeline of the 29-week trial is included in the Supplementary methods (Figure S1). During the duration of the study, the subject participated in 121 upper extremity experimental sessions, with an average of 11 sessions per week (range 5-12). Each session consisted of 100 trials during

which the subject was asked to perform motor imagery of continuous movement of the dominant right hand. The 100 trials were completed in blocks of 20 trials with several minutes of rest in between so that subject does not become fatigued. During the motor imagery tasks, the subject was instructed to think of relaxing his hand for 3 seconds, followed by thinking continuously of moving the right hand for 3 seconds during which time the ECoG activity was recorded (Figure **1C** and **D**). During study weeks 6-8, "open-loop" experiments were run where the subject was asked to perform move or rest motor imagery, but FES was not applied to the hand. Offline analysis included calculating average band power within pre-defined frequency bins (see Supplementary methods) during the "rest" and "move" periods, that were used as features for training various classifiers (bagged trees, k-nearest neighbors, linear discriminant, logistic regression, linear support vector machine, and a neural network). The purpose of training and testing various classification algorithms was to use the one which gave highest decoding accuracy and consistency, to be used for online closed-loop experiments.

In-Laboratory ECoG Power Analysis

The Activa PC+S device was configured with the following montage such that the ECoG data from channels 1 and 3 (time channels) was sampled at 200Hz, whereas channels 2 and 4 (power channels) were configured for onboard computation of the average signal power between 4-36Hz and sampled at 5Hz (see Table S3). Data from the implanted device was transmitted via the antennae to an external laptop running Matlab 2015b. Data packets were received every 400ms. For each trial, all the packets from the "rest" or "move" phase were collected to yield 3 seconds of data for each phase. The power content of both time channels for each separate experiment phase was estimated using the pspectrum function in Matlab. The frequencies were binned into 8 pre-specified segments based on typical frequency ranges commonly used to describe EEG/ECoG as: 1-8Hz, 8-12Hz, 12-18Hz, 18-26Hz, 26-35Hz, 35-45Hz, 45-70Hz, and 70-100Hz. The average signal power within each bin was computed for each of the time channels and

together with the onboard -computed power channel values used to create an 18-dimensional feature vector for each experimental phase.

TABLE S3

Channel Number	Electrode Configuration	Sample Rate (Hz)	Output Type
1	E0-E3	200	Raw Channel 1 Signal
2	E1-E2	5	Avg. Power in Channel 2 Signal between 4-36Hz
3	E8-E10	200	Raw Channel 3 Signal
4	E9-E11	5	Avg. Power in Channel 4 Signal between 4-36Hz

For channels 1 and 3, the average power in 8 pre-defined frequency ranges were used as features, whereas for channels 2 and 4 the average power between 4Hz and 36Hz yielded two additional features (Figure S3). All 18 (8 features per channel x 2 channels + 1 feature per power channel x 2 power channels) values were used as a feature vector for classifier training. Additionally, during study weeks 6-8, all bipolar combinations of the surface contacts were tested with this paradigm to determine the montage that yielded the highest power difference between rest and move signals (Figure S2). The final electrode montage is summarized in Figure **1B.** It was determined that the best performing classifier was the bagged tree classifier which was therefore chosen for online decoding for the remainder of in-laboratory experiments from weeks 9-29.

In-Laboratory Channel Montage Selection

"Open-loop" tasks which consisted of the subject performing only motor imagery of either hand rest or movement were then run with each possible electrode montage configuration and the integrated absolute difference in signal power between the rest (P_{Rest} and move (P_{Move}) phases computed in each frequency bin of interest. These trials were used to determine the electrode montage that allowed for the best discrimination between the "rest" and "move" states. As shown

in Figure S2, the configuration E0-E3 for channel 1 and E8-10 for channel 3 resulted in the largest power differences between the "rest" and "move" states and were therefore used for the remainder of the experiments.

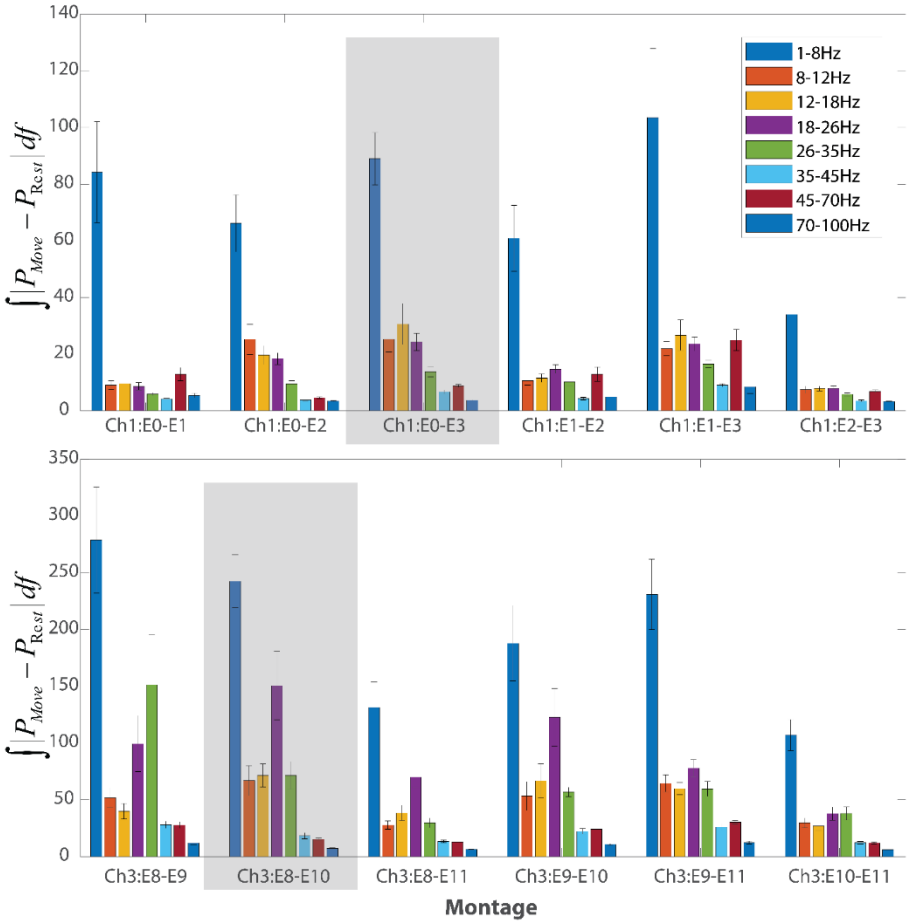


FIGURE S2 CHANNEL MONTAGE COMPARISONS. THE POWER DIFFERENCES BETWEEN THE MOVE AND REST STATES WERE INTEGRATED WITHIN A SET OF FREQUENCY BANDS LISTED IN THE LEGEND. THE MONTAGE CONFIGURATIONS HIGHLIGHTED IN GRAY WERE CHOSEN DUE TO THE ABILITY TO CAPTURE THE HIGHEST DIFFERENCES IN THE BETA BAND (18-26HZ) AND LOW GAMMA (26-35HZ).

From study weeks 9-19, "closed-loop" upper extremity experiments were conducted where the decoded motor imagery state from the online classifier was used to drive FES of the right upper extremity via an external orthosis (Bioness H200, Bioness, Valencia, CA). For these experiments, each session consisted of blocks of 20 trials. On session 1 of week 9, the first block was performed in "open loop" and used to train the online classifier for use in session 2. Subsequent sessions in closed loop were run with a classifier trained using the previous 1-5 blocks until maximum of 5

prior blocks were available for training; all subsequent blocks used the online classifier obtained from training with the prior 5 blocks of data. Figure **1D** shows average spectrograms for rest and move motor imagery across all closed-loop sessions with the final selected electrode montage which is shown in Figure **1B**.

In-Laboratory Feature Extraction and Supervised Learning Dataset Generation

Figure S3 Panel A shows a sample ECoG channel, $E(t)$, during REST (R) and MOVE (M) states during trial j. For channels 1 and 3, the power spectrum of the signals is estimated based on approximately 3 seconds worth of data for each continuous time channel sampled at 200 Hz using the pspectrum function. The average power in each bin is computed yielding 8 features for each time channel, $ch = 1,3$, and movement condition, c_j, and saved as the feature vectors $f_{ch}^{c_j}$. For Channels 2 and 4, the average power in a frequency bin centered at 20 Hz with a bin width of 16Hz (4-36Hz) is computed online within the PC+S and output from the device at a sample rate of 5Hz. The average power from Channels 2 and 4 for condition, c_j, ($f_2^{c_j}$ and $f_4^{c_j}$ respectively) are then added as additional features to those computed from Channels 1 and 3 to produce a 18-dimensional feature vector $f^{c_j} = \left[f_1^{c_j}, f_2^{c_j}, f_3^{c_j}, f_4^{c_j}\right]$ Panel B shows how for each trial j and condition, c_j, spectral features are extracted and used to construct a training set that could be used to select among different types of classifiers.

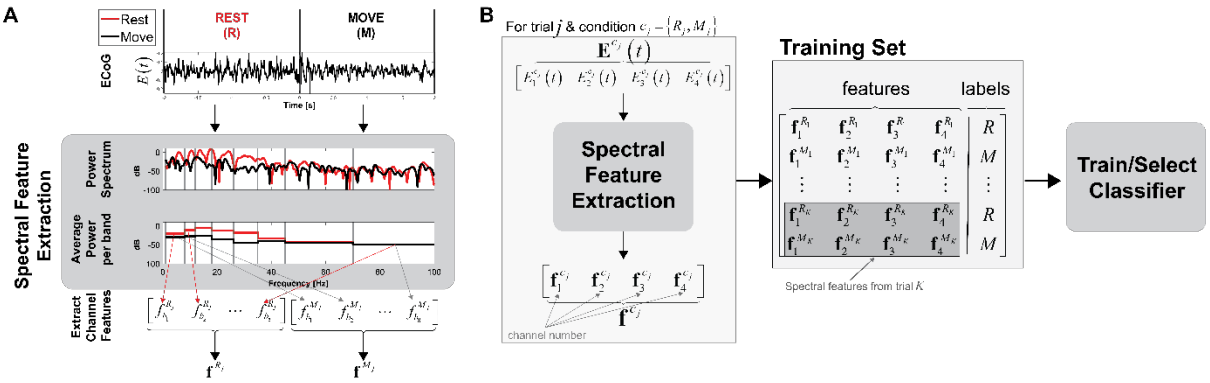


FIGURE S3 FEATURE EXTRACTION AND CLASSIFIER TRAINING DATASET. PANEL A SHOWS A SAMPLE ECOG CHANNEL, $\boldsymbol{E(t)}$, DURING REST (R) AND MOVE (M) STATES DURING TRIAL $\boldsymbol{j}$. FOR CHANNELS 1 AND 3, THE POWER SPECTRUM OF THE SIGNALS WAS ESTIMATED BASED ON APPROXIMATELY 3 SECONDS WORTH OF DATA FOR EACH CONTINUOUS TIME CHANNEL SAMPLED AT 200 HZ. THE AVERAGE POWER IN EACH BIN WAS COMPUTED YIELDING 8 FEATURES FOR EACH TIME CHANNEL, $\boldsymbol{ch = 1, 3}$, AND MOVEMENT CONDITION, $\boldsymbol{c_j}$, AND SAVED AS THE FEATURE VECTORS $\boldsymbol{f_{ch}^{c_j}}$. FOR CHANNELS 2 AND 4, THE AVERAGE POWER IN A FREQUENCY BIN CENTERED AT 20 HZ WITH A BIN WIDTH OF 16HZ (4-36HZ) WAS COMPUTED ONLINE WITHIN THE PC+S AND OUTPUT FROM THE DEVICE AT A SAMPLE RATE OF 5HZ. THE AVERAGE POWER FROM CHANNELS 2 AND 4 FOR CONDITION, $\boldsymbol{c_j}$, ($\boldsymbol{f_2^{c_j}}$ AND $\boldsymbol{f_4^{c_j}}$ RESPECTIVELY) WERE THEN ADDED AS ADDITIONAL FEATURES TO THOSE COMPUTED FROM CHANNELS 1 AND 3 TO PRODUCE A 18-DIMENSIONAL FEATURE VECTOR $\boldsymbol{f^{c_j}}$. PANEL B SHOWS HOW FOR EACH TRIAL AND CONDITION, SPECTRAL FEATURES WERE EXTRACTED AND USED TO CONSTRUCT A TRAINING SET THAT WAS USED TO SELECT AMONG DIFFERENT TYPES OF CLASSIFIERS.

Specific Classifier Parameters

All classifiers were trained in Matlab 2018b but online experiments were conducted in 2015a.

Off-line classifiers were selected as outlined in Table S4

Functional tasks

From weeks 11-19, several tasks were performed alongside the upper extremity trials to quantify any improvements in upper extremity function. Starting on week 11, whenever a correct move state was decoded and the subject was receiving FES to open and close the hand, he was asked to pick up and move a small cup (or a checker introduced from week 13) from one side of the table to the other at the center of a target. The placement accuracy was measured as a function of the distance of the cup/checker to the target. Additionally, during weeks 8-29 a modified version of the Jebsen-Taylor Hand Function Test (JHFT)[23] was performed once per week to quantify functional improvement. Passive and active range of motion was also measured each week.

Lower extremity closed-loop trials

In order to assess the ability of this class of implant to be used to control other functional movements, we sought to test the ability to trigger stepping instead of upper extremity FES. Beginning on study week 14, the subject underwent tilt table training for 1 hour per week to assist with maintaining a standing posture without significant orthostatic hypotension. Beginning on study week 20, the subject began to participate in lower extremity tasks consisting of ambulating for one hour on a robotic-assisted weight supported treadmill training device (ReoAmbulator, Motorika, Mount Laurel, NJ) 1-3 times per week. During each session, the subject would try to participate in 50 trials (with 10 trials per block). Each session was structured similarly to the upper extremity session. For each trial, the robot was configured to walk at a speed of 0.6-1.7 km per hour for 4-6 gait cycles and then stop. As the robot was slowing down, a visual cue would prompt the subject to think about moving the dominant upper extremity. If a move state was correctly decoded, the robot would be triggered to resume stepping for an additional 4-6 gait cycles (Video S1). Closed-loop lower extremity trials were conducted from study weeks 21-29.

In-Laboratory Classifiers

Figure **2A** summarizes decoding performance across all upper extremity sessions (open-loop and closed loop) for weeks 9-19 for different classifier types. For offline analysis of the closed loop experiments, a total of 80-240 trials were used with half of the data set used for training and the other half for testing. The accuracies presented represent the average of 100 monte-carlo simulations. Mean online decoding accuracy per week was 89.0% (median 88.75%, range 78-93.3%) which was not significantly different from offline performance across the 5 types of classifiers tested (Kruskall-Wallis test with Tukey-Kramer adjustment for multiple comparisons, $p>0.06$). Online decoding during weeks 9-19 remained relatively stable for upper extremity tasks across weeks as shown in Panel **2B**. Figure S4**C** summarizes decoding performance across all lower extremity gait tasks for weeks 21-29 across different classifier types. Mean online decoding accuracy per week for the lower extremity tasks was 84.15% (median 85%, range 73.3-90%).

There was no significant difference between online decoding accuracies on upper versus lower extremity tasks (two-tailed t-test, p = 0.13).

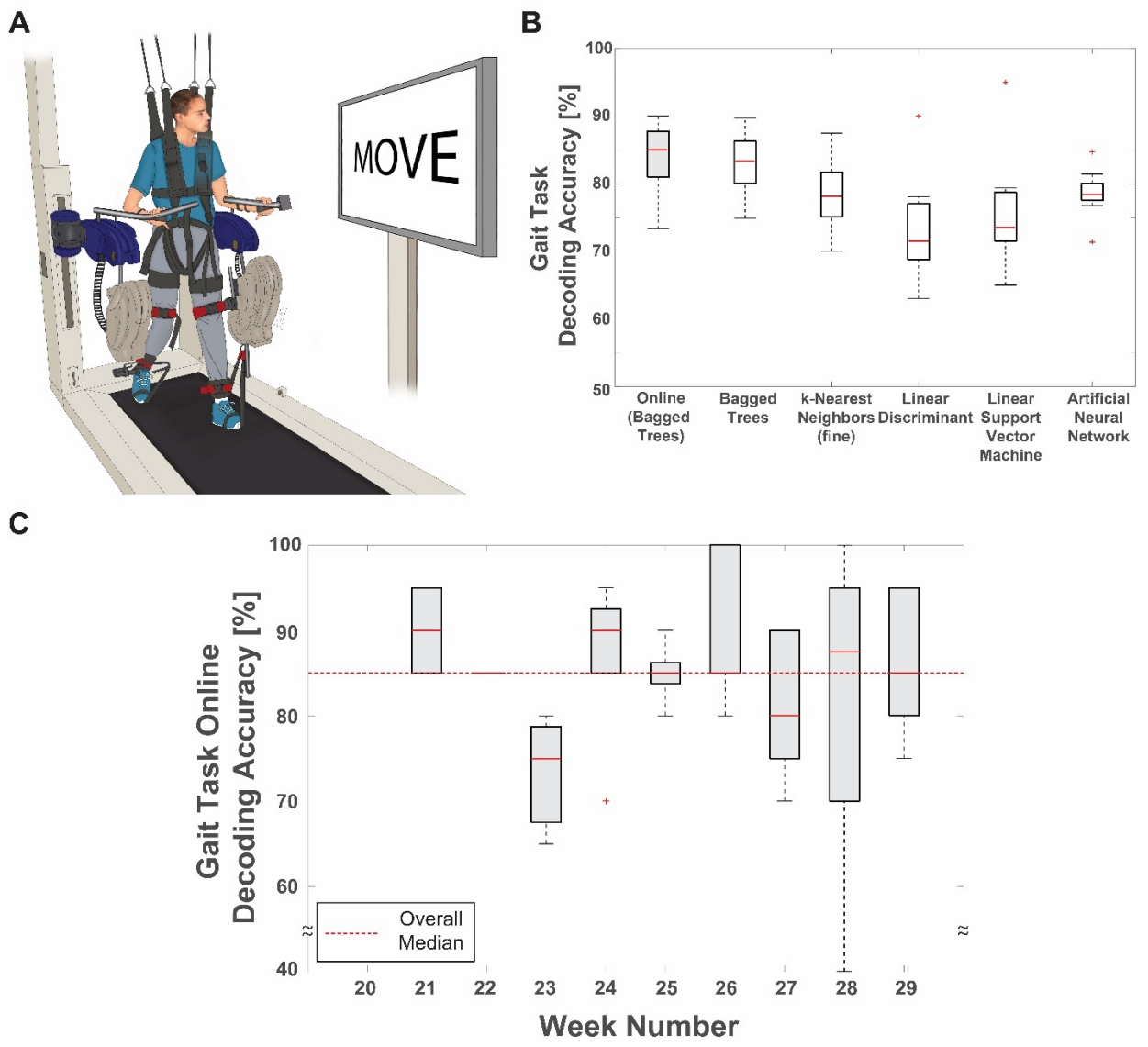


FIGURE S4 LOWER EXTREMITY DECODING PERFORMANCE. PANEL A SHOWS THE LOWER EXTREMITY AMBULATION TASK WHICH WAS PERFORMED ON A WEIGHT-SUPPORTED TREADMILL TRAINING DEVICE. WHEN A MOVE STATE WAS CORRECTLY DECODED, THE ROBOT WAS ENABLED TO CONTINUE WALKING FOR A FIXED NUMBER OF STEPS. PANEL B SHOWS THE DIFFERENT TYPES OF CLASSIFIERS TO DECODE REST/MOVE DURING THE GAIT TASK. PANEL C SHOWS THE ONLINE DECODING ACCURACY DURING THE GAIT TASK ACROSS STUDY WEEKS. ONLINE DECODING ACCURACY DURING LOWER EXTREMITY TASKS WAS SLIGHTLY MORE SENSITIVE TO SUBJECT'S ATTENTION AND THIS IS REFLECTED IN THE SLIGHTLY INCREASED VARIABILITY OF DECODING ACCURACY DURING THESE TASKS COMPARED TO THE UPPER EXTREMITY ONES.

TABLE S4 – CLASSIFIER PARAMETERS

Classifier Name	Matlab Function	Parameters
ensemble Bagged Trees	fitcensemble	MaxNumSplits=79; Method='Bag'; NumLearningCycles=30;
K nearest neighbor (kNN)	fitcknn	Distance='Euclidean', Exponent=[]; NumNeighbors=1; DistanceWeight='Equal'; Standardize='true'
linear discriminant analysis (LDA)	fitcdiscr	DiscrimType='linear'; Gamma=0; FillCoeffs='off';
support vector machine (SVM)	fitcsvm	KernelFunction ='linear'; PolynomialOrder = []; KernelScale='auto'; BoxConstraint=1; 'Standardize' = true
fully connected artificial neural network (ANN)	learnNN[2]	NumberHiddenLayers=3; ActivationFunction=tanh

Home Decoder Development

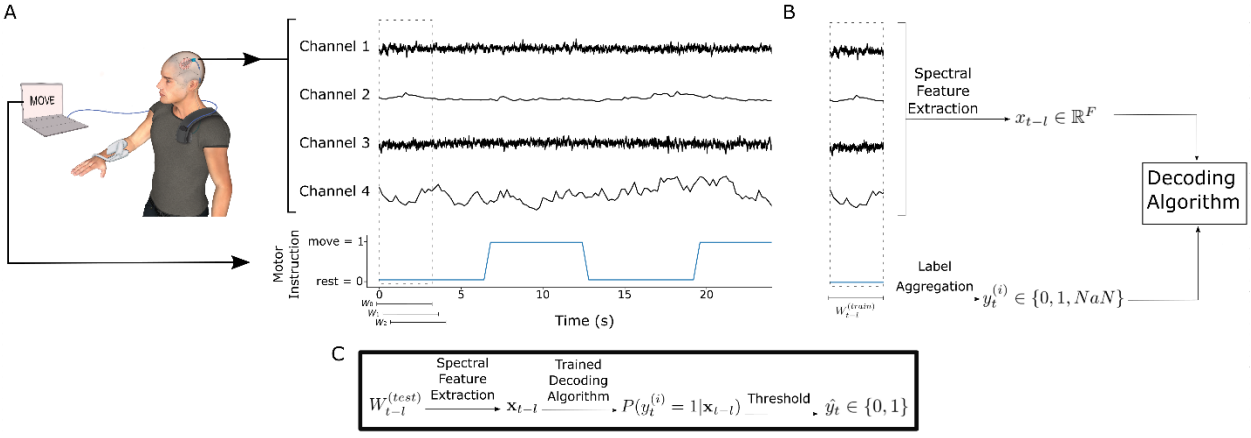


FIGURE S5 AT-HOME DECODER METHODS. PANEL A SHOWS THE RAW DATA RECORDED WITH THE DOTTED BOX REPRESENTING ONE WINDOW OF DATA. OVERLAPPING WINDOWS $\boldsymbol{w_0, w_1, w_2}$ ARE SHOWN AT THE BOTTOM WHERE EACH WINDOW WAS $\boldsymbol{w}$ SECONDS LONG WITH A WINDOW STEP OF 0.4 SECONDS. MOTOR INSTRUCTION WAS USED AS LABELS FOR DECODER TRAINING. PANEL B SHOWS THE PREPROCESSING OF THE DATA. SPECTRAL FEATURES FROM ALL FOUR CHANNELS WERE COMPUTED AND CONCATENATED, RESULTING IN FEATURE VECTOR $\mathbf{x_{t-l}}$ FOR WINDOW $\boldsymbol{w_t}$ WHERE $\boldsymbol{t}$ IS THE TIME AT THE END OF THE WINDOW AND $\boldsymbol{l}$ IS THE LAG HYPER-PARAMETER. MOTOR INSTRUCTION LABELS OVER THE WINDOW WERE AGGREGATED USING METHOD $\boldsymbol{i}$. THE RESULTING FEATURE VECTOR $\mathbf{x_{t-l}}$ AND LABEL $\boldsymbol{y_t^{(i)}}$ WERE COMPUTED FOR ALL WINDOWS $\boldsymbol{w_t}$ IN THE TRAINING SET WERE USED TO TRAIN THE DECODING ALGORITHM. HYPER-PARAMETERS $\{\boldsymbol{w, l, i, \mathbf{decoding\ algorithm}}\}$ WERE SELECTED VIA CROSS-VALIDATION (TABLE S5). PANEL C SHOWS A FLOW DIAGRAM OF ONLINE DECODING OF MOTOR STATE AS APPLIED TO THE OPEN-LOOP AND CLOSED-LOOP AT-HOME DATA.

At-Home Feature Extraction

Feature extraction and decoder development hyper-parameters were selected by leave-one-out cross-validation over the 33 trials in the at-home training data set. The first hyper-parameter was window size for the ECoG spectral estimate, $w \in \{0.8, 1.2, 1.6, 2.0, 2.4, 2.8, 3.2\}$ measured in seconds. Overlapping windows, with a step of 0.4 seconds, resulted in N windows per trial where N varied with window duration w. As a result of windowing, motor instruction labels were aggregated from each sample in the window, where 1 indicates motor intent and 0 indicates rest. In order to build a decoding model robust to transitions from a move to rest state, windows with data collected during both move and rest states were not removed for training. Three possible methods for aggregating windows of labels were considered during cross-validation (Figure **S6**). For each window, vector $\mathbf{y} = \{y^{(1)}, y^{(2)}, y^{(3)}\}$ of labels was computed where

- $y^{(1)} \in \{1,0\}$ is the last label of the window.
- $y^{(2)} \in \{1,0\}$ is the majority of labels in the window.
- $y^{(3)} \in \{1,0,NaN\}$ is 1 or 0 if there is unanimous consensus within the window (all 1 or all 0), or NaN if there are both move and rest labels in the window. This aggregation is motivated by aiming to provide the decoder with the highest quality data. NaN-labelled windows were not used for supervised learning.

For ECoG channels 1 and 3, the PSD of each window was computed using the multitaper method in the MNE python package[3] with normalized half-bandwidth $\mathrm{T} = 3$, and a bandwidth of $b = \frac{2\times\mathrm{T}}{0.4\times\mathrm{w}}$ [4]. $2\mathrm{T} - 1 = 5$ tapers were used and an adaptive weighting routine was used to combine estimates of different tapers[5]. Spectral power was estimated for a frequency range between 0-100Hz and was converted to decibels. Spectral estimation for channels 1 and 3 resulted in $x^{(ch\,1)} \in \mathbb{R}^{F_{psd}}$

and $x^{(ch\,3)} \in \mathbb{R}^{F_{psd}}$, respectively and where F_{psd} is the length of the PSD. For channels 2 and 4, median power $\tilde{x} \in \mathbb{R}_+$ for each window was calculated. Spectral estimates from all channels were aggregated into one spectral feature vector $\mathbf{x} \in \mathbb{R}^F$ for each window where $F = 2 \times F_{psd} + 2$

At-Home Decoder Architecture

Three decoding model architectures were considered during cross-validation. Each decoding model used a fixed window length and labeling method.

Hidden Markov Model decoder

In order to incorporate the temporal dynamics of switching between move and rest states, we used a Hidden Markov Model, a state-space model that has been used to describe time series data in a wide variety of fields[6]. It assumes an M-state system has $\{q_0, q_1, \ldots, q_{M-1}\}$ discrete latent states which evolve over time, driven by a first-order, ergodic Markov chain resulting in a sequence of N states $Z = (z_0, z_1, \ldots, z_{N-1})$. Observations of the system $(\boldsymbol{x}_n)$ are distributed according to state-specific Gaussian emission distributions $\mathbf{B} = \{\mathbf{b}_{\boldsymbol{m}}\}$ where $\mathbf{b}_{\boldsymbol{m}} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{m}}, \boldsymbol{\Sigma}_{\boldsymbol{m}})$ for each state q_m with mean $\boldsymbol{\mu}_{\boldsymbol{m}}$ and covariance $\boldsymbol{\Sigma}_{\boldsymbol{m}}$, which was constrained to be diagonal so that training would be more computationally tractable. The Markov chain transition matrix is $A = \{a_{ij}\}$ where $a_{ij} = Pr(z_{n+1} = q_j | z_n = q_i)$. The initial state of the system is drawn from the discrete initial state distribution $\boldsymbol{\pi}$. The entire HMM is fully parameterized by $\lambda = (A, \mathbf{B}, \boldsymbol{\pi})$. In the model system, the state of the system at each discrete time n is based on the Markov chain transition probabilities and an observed feature is generated according to current state z_n resulting in a sequence of observations $\mathbf{X} = (\mathbf{x_0}, \mathbf{x_1}, \ldots \mathbf{x_{N-1}})$**.** [6,7]

Each HMM was trained with the Baum-Welch algorithm with random parameter initialization and a maximum of 10 iterations using the hmmlearn python package [8]. For prediction of latent state, the forward algorithm was used in order to estimate the probability of being in each state at each time n. The normalized forward algorithm computes the probability of being in each latent state

q_i, at time n, and is defined as $\alpha_n(q_i) = Pr(z_n = q_i \mid x_0, \dots, x_n, \lambda) = \hat{x}^{(i)}$ resulting in $\hat{\mathbf{x}}_\mathbf{n} \in [0,1]^M$ where M is the number of HMM states and $\sum_{i=1}^{M} \hat{x}^{(i)} = 1$. We let the number of states in the HMM vary with $M \in \{3,5,7\}$ because although our target was binary, a number of different states could be reflected in the neural signal related to different aspects hand grasp initiation and termination. M was selected in cross-validation (Figure S6, Table S5)

Logistic regression (LR) is used to map HMM state probabilities $\hat{\mathbf{x}}_\mathbf{n}$ to binary move/rest targets. For each window beginning at time point n that did not have a NaN label, the feature vector is accompanied by a label $y_n \in \{0,1\}$ indicating whether the window corresponds with rest (0) or motor intent (1). NaN labeled windows were dropped for fitting LR parameters. A LR model is parameterized by a vector $\boldsymbol{\beta} \in \mathbb{R}^{M+1} = [\beta_0, \beta_1, \dots, \beta_M]^T$. For a given parameterization $\boldsymbol{\beta}$, the LR estimated likelihoods of motor intent for window starting at time point n are given by:

$$Pr(\text{Motor Intent} \mid \hat{\mathbf{x}}_n; \boldsymbol{\beta}) = \frac{\exp \boldsymbol{\beta}^T \tilde{\mathbf{x}}_n}{1 + \exp \boldsymbol{\beta}^T \tilde{\mathbf{x}}_n}$$

$$Pr(\text{Rest} \mid \hat{\mathbf{x}}_n; \boldsymbol{\beta}) = \frac{1}{1 + \exp \boldsymbol{\beta}^T \tilde{\mathbf{x}}_n}$$

Where $\tilde{\mathbf{x}}_n = [1, \hat{\mathbf{x}}_\boldsymbol{n}]^T$ is the HMM state probability vector with a one prepended, enabling β_0 to serve as a constant offset. Thus, fitting an LR model entails finding the parameters that maximize the elastic net regularized log-likelihood of the labels corresponding to the training data:

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \frac{1-\rho}{2} \|\boldsymbol{\beta}\|_2^2 + \|\boldsymbol{\beta}\|_\mathbf{1} + \sum_{n=1}^{N} -\log \Pr\left(y_n | \mathbf{x}_n; \boldsymbol{\beta}\right)$$

Where $\|\boldsymbol{\beta}\|_2^2 = \boldsymbol{\beta}^T \boldsymbol{\beta}$ and $\boldsymbol{\rho} = \mathbf{0.5}$. The LR parameter vector $\widehat{\boldsymbol{\beta}}$ was computed using scikit-learn with elasticnet regularization and the SAGA solver. [8]

Linear Discriminant Analysis Decoder

The second decoding model architecture was a Linear Discriminant Analysis (LDA) decoder. Since the observations to the HMM decoder were the derived spectral features, the latent states fitted by the HMM are not guaranteed to be related to motor intent. LDA is a supervised classification technique that separates groups of labelled data by maximizing the ratio of between-class to within-class separability.[9] This yields a *linear discriminant* vector which maximally separates the two classes.

First, the training data $\mathbf{X} \in \mathbb{R}^{F \times N}$ were divided by motor intent label into the subset $\mathbf{X}^{(0)} \in \mathbb{R}^{F \times N_0}$ for data labelled as a rest state, $\mathbf{X}^{(1)} \in \mathbb{R}^{F \times N_1}$ for data labeled as a move state, and $\mathbf{X}^{(NaN)} \in \mathbb{R}^{F \times N_{NaN}}$ for data without a label (see Windowing). $\mathbf{X}^{(NaN)}$ was not used for parameter fitting as LDA is a supervised learning method which necessitates labelled data. For $j \in \{0,1\}$ and $N^* = N_0 + N_1$, let $\boldsymbol{\mu}_j$ be the sample mean of $\mathbf{X}^{(j)}$ and define the *scatter matrix* as the unnormalized sample covariance matrix $\mathbf{M}_j = \sum_{n \in N^*} \left((\mathbf{x}_n^{(j)} - \boldsymbol{\mu}_j) \left(\mathbf{x}_n^{(j)} - \boldsymbol{\mu}_j \right)^T \right)$. *Within-class* scatter matrix is defined as $\mathbf{M}_W = \mathbf{M}_0 + \mathbf{M}_1$ and the *between-class* scatter matrix was defined as $\mathbf{M}_B = (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^T$. Thus, the linear discriminant $\mathbf{v}^* \in \mathbb{R}^F$ is found as the solution to:

$$\mathrm{v}^* = \underset{\mathrm{v}}{\operatorname{argmax}} \frac{\mathrm{v}^T \mathrm{M}_B \mathrm{v}}{\mathrm{v}^T \mathrm{M}_W \mathrm{v}}$$

Which maximizes the variance between classes while minimizing the variance within classes. LDA models were implemented using the scikit-learn Python package. [8]

LDA-HMM Decoder

The third decoder architecture was a combination of the described LDA and HMM decoders. The LDA parameters were fitted as described above, but rather than using LDA directly for classification, the spectral feature vector $\mathbf{x}_n$ was used to generate the LDA scores $\mathbf{x}_n^* = \mathbf{x}_n^T \mathbf{v}^*$ for

every window n. The sequence of LDA scores $\mathbf{X}^* = \left(\mathbf{x}_0^*, \mathbf{x}_1^*, \ldots, \mathbf{x}_{N-1}^*\right)$ was then used as the observations for a 2-state HMM decoder which generated prediction of motor state as described above.

Lagged decoding

Introducing a time lag in decoding between neural features and the label can contribute to increases in decoder performance in motor BCIs.[10,11] Because packets of data are transmitted every 0.4 seconds from the recording device, the values considered for a lag were multiples of 0.4 seconds. Parameter $l \in \{0.8, 0.4, 0.0\}$ is the number of seconds feature vector $\mathbf{x}$ precedes the associated label y.

Cross-validation

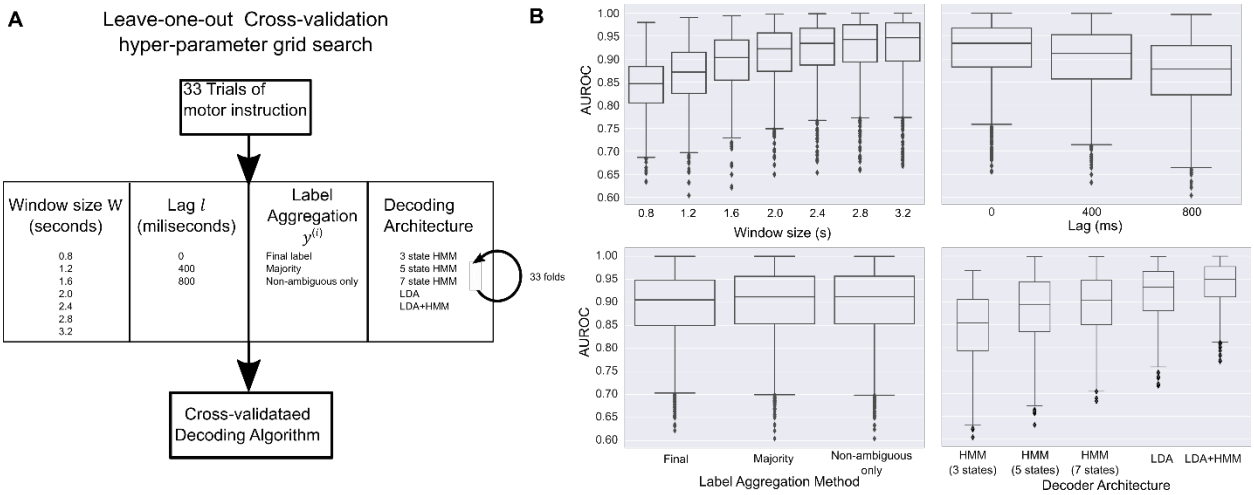


FIGURE S6 CROSS-VALIDATION OVERVIEW **A)** GRID SEARCH FOR LEAVE-ONE-TRIAL-OUT CROSS-VALIDATION DEPICTED FOR 33 TRIALS OF MOTOR INSTRUCTION. SELECTED HYPER-PARAMETERS ARE SUMMARIZED IN TABLE S5. **B)** IMPACT OF EACH HYPER-PARAMETER PLOTTED OVER ALL OTHER HYPER-PARAMETERIZATIONS. WINDOW SIZE, LAG, AND DECODER ARCHITECTURE HAD LARGE IMPACTS ON PERFORMANCE, BUT LABEL AGGREGATION METHOD HAD A SIMILAR DISTRIBUTION OF PERFORMANCE OVER ALL OTHER HYPER-PARAMETERIZATIONS.

In order to select the optimal set of hyper-parameters, we used leave-one-out cross-validation over the 33 trials in the train dataset. The grid search for each parameter was

- **Window size:** $w \in \{0.8, 1.2, 1.6, 2.0, 2.4, 2.8, 3.2\}$ where w is the duration in seconds of the window of neural data used for each prediction.

- **Label Scheme:** $y^{(i)} \in \mathbf{y}$ where y is one of the three labeling schemes defined above
- **Decoder Architecture:** A total of 5 decoder architectures were considered. Three different HMM decodes were considered, one for each value of $M \in \{3, 5, 7\}$**.** The above-described LDA decoder and LDA-HMM decoder were also considered.
- **Lag:** $l \in \{0.8, 0.4, 0.0\}$ where l is seconds between feature vector x and target label y

We compared cross-validated models using median area under the receiver-operator characteristic curve (AUC) for each fold. In order to compute the AUC for each fold, we chose to use the non-ambiguous labeling scheme $y^{(3)}$ to bias the hyperparameters for reliability in decoding. Impact of each hyperparameter on performance during cross-validation is summarized in Figure S6B and the set of hyperparameters that were selected can be found in Table S5.

Hyperparameter	Selected Value
Window size, w (seconds)	3.2
Label Scheme, y	$y^{(3)}$
Decoder Architecture	LDA-HMM
Lag l (seconds)	0.0

TABLE S5 FINAL DECODER HYPER-PARAMETERS SELECTED VIA LEAVE-ONE-OUT CROSS-VALIDATION

Threshold setting

The decoded motor intent was calculated by thresholding the predicted probability of motor intent for the final trained model. The receiver-operator characteristic (ROC) curve was calculated using the training dataset to calculate Youden's J score across different threshold. Youden's J score balances the sensitivity and specificity of the decoder for a specific threshold. The optimal threshold was selected by maximizing Youden's J score, resulting in a value of 0.969.

Artifact Detection

The Medtronic PC+S recording device included quality control protocols which resulted in periodic electrical artifacts occurring approximately every 10 minutes. When artifacts were detected,

decoding was paused for 0.8 seconds until the artifact passed, during which the previously decoded motor state was maintained.

Functional Task Accuracy

A relative distance score, r_i, was assigned to each target based on the distance of the object from the center of the target as shown in Figure 3A. A placement score for the ith object placement was computed as

$$S_i = \frac{1}{1 + r_i}$$

so that a higher score corresponded to objects placed closer to the center of the target. The functional task was repeated a total of 20 times during each block leading to a maximum score $S_{t_{MAX}} = 20$ per block. The subject's accuracy was determined then as the sum of all the individual trial scores, S_t, divided by maximum block score:

$$accuracy = \frac{S_t}{S_{t_{MAX}}} = \frac{\sum_{i=1}^{20} S_i}{S_{t_{MAX}}}$$

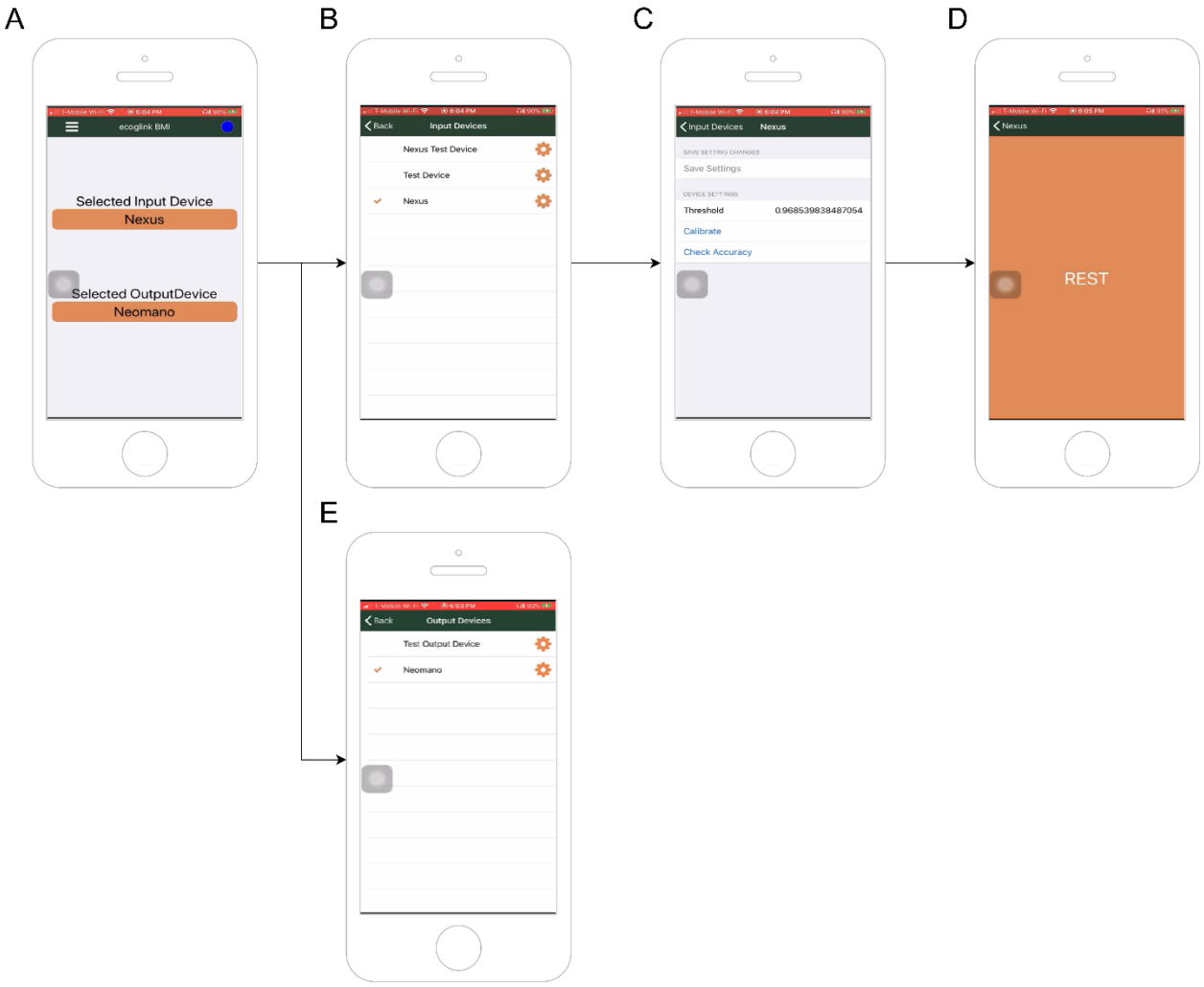
The subject showed improvement in the mean ($\pm$ std. error) accuracy of placing a small cup, 60.1% $\pm$ 7.8% at week 11 versus 82.8% $\pm$ 4.7% at week 19 (p=0.03) or a checker (64.5% $\pm$ 7.3% at week 13 versus 88.8% $\pm$ 4.8% at week 19, p=0.03) at the center of a target as summarized in Figure **3A** and **3B**.

Functional improvement was quantified as the reduction in the average time taken to perform specific components of the JHFT (Figure **3C**). Significant improvements were observed in lifting small objects, lifting light cans, and lifting heavy cans through orthotic-assisted tasks. Along with a trend towards improvement in writing speed (32.3s to 26.4s, p=0.15), clarity of the handwriting

also improved throughout the course of the study (Figure **3D**). Further, pinch force increased from 1lb to 3lb within 10 weeks.

Phone-based User Application Development

A custom-made mobile application was designed allowing the subject to interact with and modify settings of the BCI. The home screen (Figure **S7A**) displays the currently connected and selected devices in use by the subject. The application was designed by generalizing devices that could connect to the system, thus providing a method to select from a list of connected input (Figure **S7B**) and output (Figure **S7E**) devices. These list views provided links to device-specific settings such as the decoder threshold linked to incoming data from the Activa PC+S (Figure **S7C**). The settings also allowed the subject to initiate data collection sessions to assess the accuracy of the current decoder (Figure **S7D**).



407

408 **FIGURE S7** FLOW DIAGRAM OF THE MOBILE APPLICATION USED BY THE SUBJECT FOR AT-HOME INTERACTION AND
409 ADJUSTMENT OF THE BCI

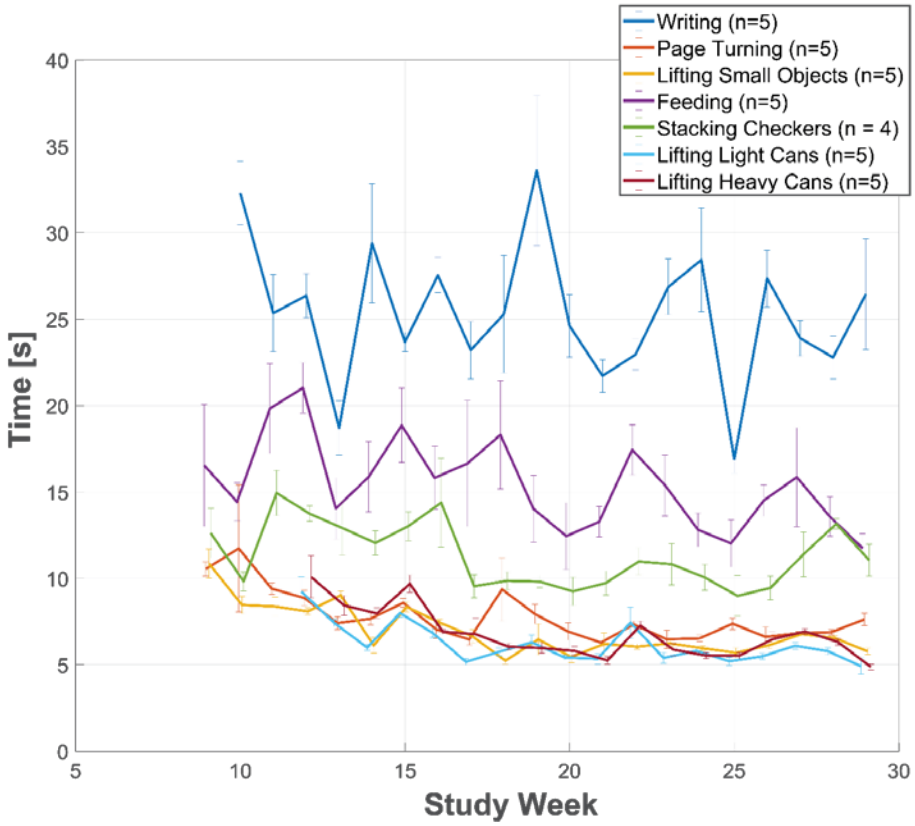


FIGURE S8 JEBSEN HAND FUNCTION TEST OVER THE STUDY COURSE. ERROR BARS REPRESENT STANDARD ERROR FROM THE MEAN. FIGURE 3C COMPARES THE FIRST AND LAST SESSIONS.

Clinical Assessments

The subject underwent weekly interviews to assess for adverse events and was also surveyed for changes in self-perceived functional independence. Changes in health status were assessed with the MOS 36-item shot form health survey (SF-36).[12] Perceived changes in functional independence were assessed with the Spinal Cord Independence Measure (SCIM) version III[13] which ranges from 0 to 100 and higher score indicate increased independence. Detailed neurological evaluation for documentation of level and severity of SCI was conducted monthly according to the ISNCSCI. [14]

While there was no change in ISNCSCI ASIA impairment scale from a C5 motor level, there was an unexpected slight increase in the motor zone of partial preservation (defined as the myotomes below the level of injury with residual innervation) on the left from C6 to C8. Additionally, after study week 23, the subject began to be able to extend his right thumb volitionally with motor strength 2/5 in the absence of the FES orthosis. There was no change in the SCIM from a baseline score of 26. The SF-36 indicated a 32.5% improvement in pain, a 5% increase in energy, and an 8% decrease in emotional well-being

Detailed neurological evaluation for documentation of level and severity of SCI was conducted monthly according to the ISNCSCI. Figure S9 and S10 summarize the results of the ISNCSCI obtained during the week 1 and week 29 visits. Diagrams generated using the European Multicenter Study about Spinal Cord Injury (EMSC) ISNCSCI calculator.[15]

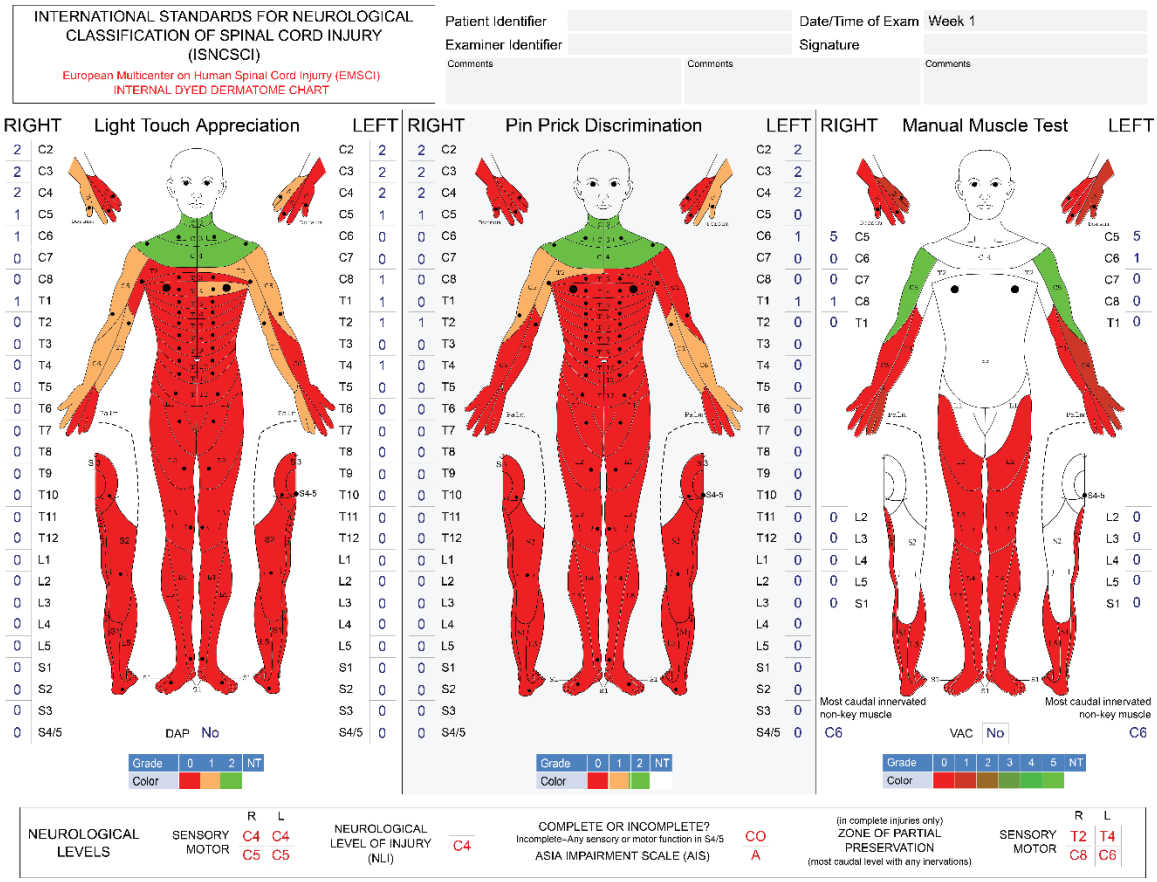


FIGURE S9 ISNCSCI EXAM ON INITIAL VISIT – STUDY WEEK 1. NOTE THE C5 MOTOR LEVEL AND C4 NEUROLOGICAL LEVEL OF INJURY (NLI) DUE TO DIMINISHED SENSATION IN THE C5 DISTRIBUTION. THE ZONE OF PARTIAL PRESERVATION FOR THE RIGHT/LEFT IS T2/T4 FOR SENSORY FUNCTION AND C8/C6 FOR MOTOR.

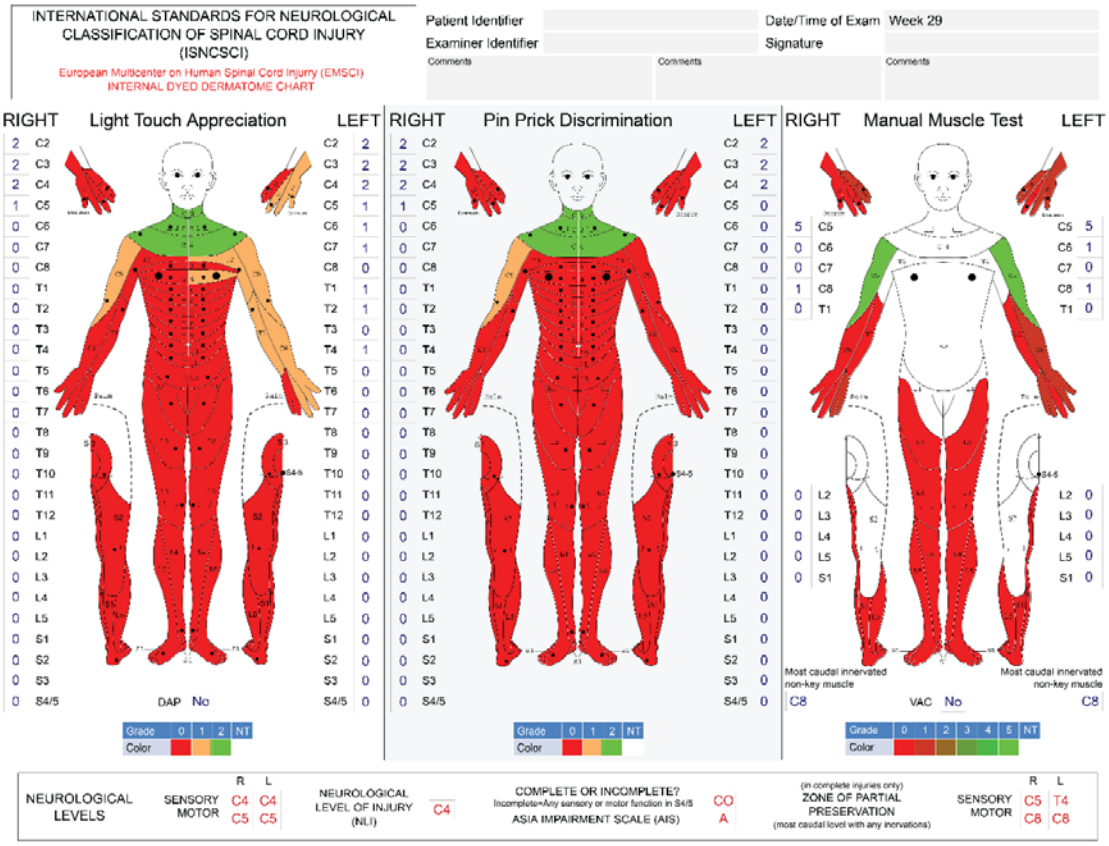


FIGURE S10 ISNCSCI EXAM ON FINAL VISIT – WEEK 29. NOTE THAT WHILE THE MOTOR LEVEL REMAINS C5 AND THE NLI C4, THERE IS A SLIGHT INCREASED ZONE OF PARTIAL PRESERVATION WITH IS NOW C8 BILATERALLY COMPARE TO C8 ON THE RIGHT AND C6 ON THE LEFT DURING WEEK ONE. ADDITIONALLY, THE SUBJECT GAINED THE ABILITY TO SLOWLY EXTEND HIS THUMB ON COMMAND ON THE RIGHT SIDE WHICH IS NOTED AS THE MOST CAUDAL INNERVATED NON-KEY MUSCLE AS C8 ON THE RIGHT PANEL.

Changes in health status were assessed with the MOS 36-item shot form health survey (SF-36).[12] Comparisons of SF-36 scores between initial and final study visit are summarized in Table S5. Perceived changes in functional independence were assessed with the Spinal Cord Independence Measure (SCIM)[13] version III which ranges from 0 to 100 and higher score indicate increased independence. Changes are in SCIM are summarized in Table S6.

There was an 8% decrease in emotional well-being in the SF36 changes in responses to the following questions:

1) "During the past 4 weeks, have you been a happy person?" changing from "all of the time" to "most of the time."

2) "During the past 4 weeks, have you felt calm and peaceful?" changing from "all of the time" to most of the time."

The 32.5% improvement in the pain score was driven by changes in responses to the following questions:

1) "How much bodily pain have you had in the past 4 weeks?" changing from "mild" to "very mild."
2) "During the past 4 weeks, how much did pain interfere with your normal work?" – change from "moderately" to "a little bit."

The 5% improvement is energy/fatigue was driven by the response to the question "During the last 4 weeks, did you feel worn out?" changing from "a little bit of the time" to "none of the time."

TABLE S6 SF36 SCORES CHANGES. SCORES CALCULATED USING [12,16]

	Laboratory Trials			*Home Trials*		
Category	**Initial Visit**	**Final Visit**	**Change**	**Initial Visit**	**Final Visit**	**Change**
Physical Functioning	0%	0%	**0%**	0%	0%	**0%**
Role Limitations Due to Physical Health	100%	100%	**0%**	100%	100%	**0%**
Role Limitations Due to Emotional Problems	100%	100%	**0%**	100%	100%	**0%**
Energy/Fatigue	80%	85%	**5%**	90%	90%	**0%**
Emotional Well-Being	100%	92%	**-8%**	92%	92%	**0%**
Social Functioning	100%	100%	**0%**	100%	100%	**0%**
Pain	67.5%	100%	**32.5%**	100%	77.5%	**-22.5%**
General Health	85%	85%	**0%**	85%	90%	**5%**
Health Change	75%	75%	**0%**	75%	75%	**0%**

TABLE S7 SPINAL CORD INDEPENDENCE MEASURE.

	Laboratory Trials			*Home trials*		
	Initial Visit	**Final Visit**	**Change**	**Initial Visit**	**Final Visit**	**Change**
Self-Care	2	2	**0**	2	3	**1**
Respiration and Sphincter Management	21	21	**0**	21	21	**0**
Mobility	3	3	**0**	3	3	**0**
Total	26	26	**0**	26	27	**1**

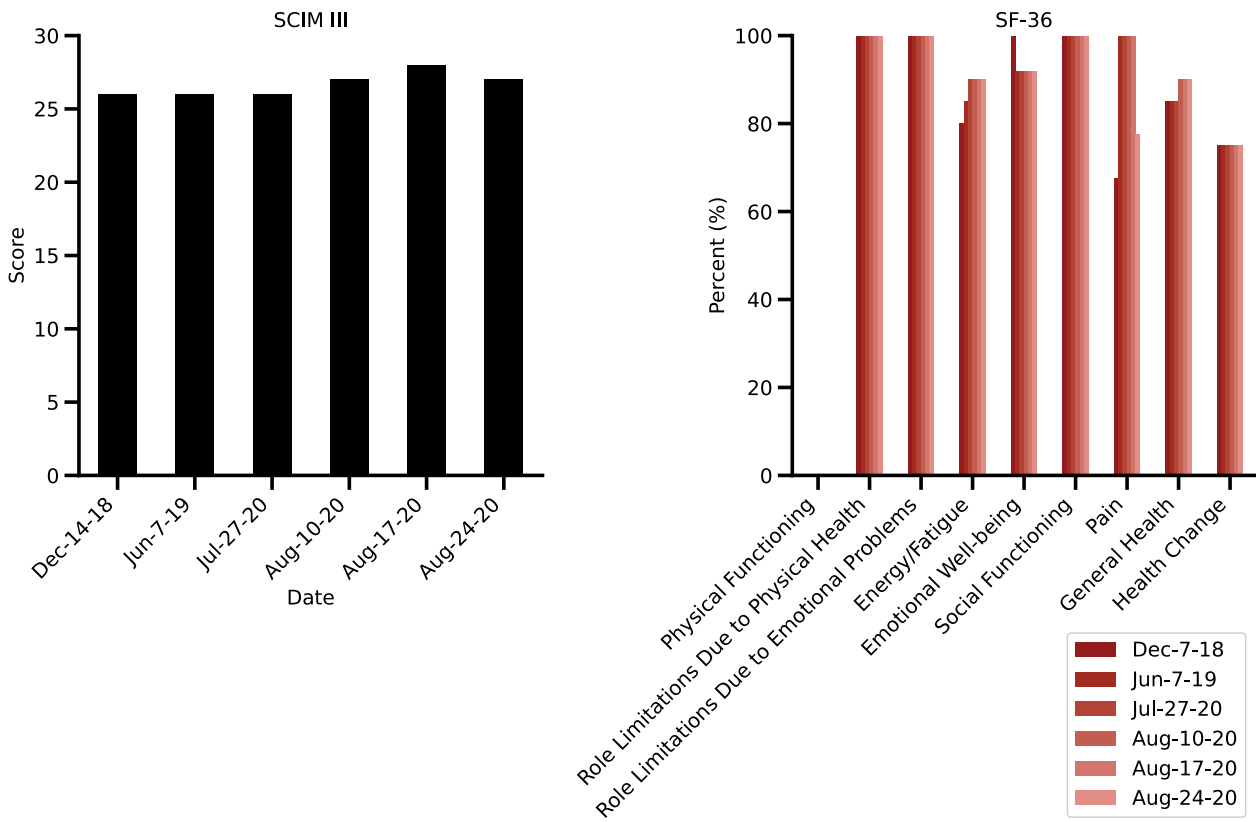


FIGURE S11 SCIM III AND SF-36. CHANGES IN SCORES FOR THE SCIM III (LEFT) AND SF-36 (RIGHT) OVER THE COURSE OF THE STUDY PERIOD. THE FIRST TWO TIME POINTS WERE TAKEN BEFORE AND AFTER THE LABORATORY STUDY PERIOD. THE FINAL 4 TIME POINTS WERE TAKEN THROUGHOUT THE STUDY AT HOME.

References

1. Gant K, Guerra S, Zimmerman L, Parks BA, Prins NW, Prasad A. EEG-controlled functional electrical stimulation for hand opening and closing in chronic complete cervical spinal cord injury. Biomedical Physics & Engineering Express 2018;4:065005.
2. Tshitoyan V. Simple Neural Network. 1.1 ed. GitHub: GitHub; 2019.
3. Gramfort A, Luessi M, Larson E, et al. MEG and EEG data analysis with MNE-Python. Front Neurosci 2013;7:267.
4. Babadi B, Brown EN. A review of multitaper spectral analysis. IEEE Trans Biomed Eng 2014;61:1555-64.
5. Thomson DJ. Spectrum estimation and harmonic analysis. Proceedings of the IEEE 1982;70:1055-96.
6. Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006.
7. Rabiner LR. A Tutorial on Hidden Markov-Models and Selected Applications in Speech Recognition. Proceedings of the Ieee 1989;77:257-86.
8. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res 2011;12:2825-30.
9. Tharwat A, Gaber T, Ibrahim A, Hassanien AE. Linear discriminant analysis: A detailed tutorial. Ai Communications 2017;30:169-90.

10. Carmena JM, Lebedev MA, Crist RE, et al. Learning to control a brain-machine interface for reaching and grasping by primates. PLoS Biol 2003;1:E42.
11. Presacco A, Goodman R, Forrester L, Contreras-Vidal JL. Neural decoding of treadmill walking from noninvasive electroencephalographic signals. J Neurophysiol 2011;106:1875-87.
12. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
13. Itzkovich M, Gelernter I, Biering-Sorensen F, et al. The Spinal Cord Independence Measure (SCIM) version III: reliability and validity in a multi-center international study. Disabil Rehabil 2007;29:1926-33.
14. Kirshblum SC, Burns SP, Biering-Sorensen F, et al. International standards for neurological classification of spinal cord injury (revised 2011). J Spinal Cord Med 2011;34:535-46.
15. ISNCSCI Calculator. European Multicenter Study about Spinal Cord Injury, 2019. (Accessed 06/18/2019, 2019, at http://ais.emsci.org/.)
16. Free online SF-36 score calculator. OrthoToolKit, 2019. (Accessed 7/2/2019, 2019, at https://www.orthotoolkit.com/sf-36/.)