Decoding continuous kinetic information of grasp from stereo-electroencephalographic (SEEG) recordings

Journal of Neural Engineering

PAPER

Decoding continuous kinetic information of grasp from stereo-electroencephalographic (SEEG) recordings

Xiaolong Wu[1,4], Guangye Li[2,4], Shize Jiang[3], Scott Wellington[1], Shengjie Liu[2], Zehan Wu[3], Benjamin Metcalfe[1], Liang Chen[3,*] and Dingguo Zhang[1,*]

[1] Centre for Autonomous Robotics (CENTAUR), Department of Electronic and Electrical Engineering, University of Bath, Bath, United Kingdom
[2] State Key Laboratory of Mechanical Systems and Vibrations, Institute of Robotics, Shanghai Jiao Tong University, Shanghai, People's Republic of China
[3] Department of Neurosurgery of Huashan Hospital, Fudan University, Shanghai, People's Republic of China
[4] Contribute to this paper equally and should be considered as co-first authors.
* Authors to whom any correspondence should be addressed.

E-mail: vhschenliang@fudan.edu.cn **and** d.zhang@bath.ac.uk

Keywords: stereo-electroencephalography (SEEG), brain–computer interface (BCI), kinetic decoding, grasp force, deep learning

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
10 October 2021

REVISED
4 April 2022

ACCEPTED FOR PUBLICATION
8 April 2022

PUBLISHED
21 April 2022

Abstract

Objective. Brain–computer interfaces (BCIs) have the potential to bypass damaged neural pathways and restore functionality lost due to injury or disease. Approaches to decoding kinematic information are well documented; however, the decoding of kinetic information has received less attention. Additionally, the possibility of using stereo-electroencephalography (SEEG) for kinetic decoding during hand grasping tasks is still largely unknown. Thus, the objective of this paper is to demonstrate kinetic parameter decoding using SEEG in patients performing a grasping task with two different force levels under two different ascending rates. *Approach.* Temporal-spectral representations were studied to investigate frequency modulation under different force tasks. Then, force amplitude was decoded from SEEG recordings using multiple decoders, including a linear model, a partial least squares model, an unscented Kalman filter, and three deep learning models (shallow convolutional neural network, deep convolutional neural network and the proposed CNN+RNN neural network). *Main results.* The current study showed that: (a) for some channel, both low-frequency modulation (event-related desynchronization (ERD)) and high-frequency modulation (event-related synchronization) were sustained during prolonged force holding periods; (b) continuously changing grasp force can be decoded from the SEEG signals; (c) the novel CNN+RNN deep learning model achieved the best decoding performance, with the predicted force magnitude closely aligned to the ground truth under different force amplitudes and changing rates. *Significance.* This work verified the possibility of decoding continuously changing grasp force using SEEG recordings. The result presented in this study demonstrated the potential of SEEG recordings for future BCI application.

1. Introduction

Brain–computer interfaces (BCIs) acquire, analyze, and translate electrical signals originating from the brain into commands that can be used to drive external actuators, independent of the user's physical abilities. To achieve this, both non-invasive and invasive recording techniques have been developed, including scalp electroencephalography (EEG) (Wolpaw *et al* 2002), electrocorticography (ECoG) (Schalk and Leuthardt 2011), stereo-electroencephalography (SEEG) (Herff *et al* 2020), and single neuron spiking recording (Orsborn *et al* 2015). However, the decision to employ either an invasive or noninvasive recording method is non-trivial, as both have inherent benefits and constraints that must be considered. EEG records signals directly from the surface of the scalp, and therefore

have greater convenience and safety. However, signals recorded using scalp EEG show lower spatial selectivity and signal-to-noise ratio than invasive approaches in which the electrodes are placed directly on or in the brain. For example, SEEG depth electrodes can record neural activity from both shallow and deep brain structures, such as the hippocampus or basal ganglia, which are generally not accessible with approaches such as ECoG (attached to the surface of the cortex subdurally or epidurally) or intracortical arrays (1–1.5 mm beneath the cortex). A number of studies have tested the preliminary performance of SEEG signals for BCI applications, such as kinematic decoding (Li *et al* 2017, Wang *et al* 2020), kinetic decoding (Tan *et al* 2013), spelling system using P300 (Huang *et al* 2021) and speech decoding (Meng *et al* 2021).

Of particular interest for the BCIs application is the decoding of both continuous kinematic (e.g. trajectories, velocity, and position) and kinetic (e.g. force) parameters. Extensive studies have been undertaken in kinematic decoding; for example, ECoG has been used to decode continuous finger joint angles and hand positions in both humans and rhesus macaque monkeys (Flint *et al* 2017, Farrokhi and Erfanian 2020), as well as being used in human to achieve satisfactory reconstruction of hand and arm trajectories (Chen *et al* 2013, Nakanishi *et al* 2013). Despite the progress made in decoding kinematic information, there have been relatively few trials demonstrating the decoding of kinetic information, especially for continuously changing forces. Among these few, one EEG-based study reported that the lower beta-band frequencies in the centro-parietal region were found to reflect an object's shape and size during the grasp pre-shaping stage, whereas during the grasp finalization and holding stages, the mu frequency band in the contralateral parietal region reflected muscle activity (Sburlea and Müller-Putz 2018). Cassim *et al* demonstrated event-related synchronization (ERS) and desynchronization (ERD) modulation in the mu and beta bands during wrist extension by studying recordings from eight healthy participants using surface EEG; however, the power in these bands returned to baseline within 4–5 s even when the wrist position was maintained (Cassim *et al* 2000). In another EEG study, good classification performance was achieved between three hand grasp types within a reach-and-grasp paradigm (Schwarz *et al* 2017), achieving a binary classification accuracy of 72.4% between grasp types. For invasive methods, ECoG has been applied to the left primary motor cortex (M1) in Japanese macaques to decode hand muscle activation (Shin *et al* 2012). In humans, Flint *et al* used ECoG to decode continuous isometric pinch force from 10 human participants (Flint *et al* 2014). The predicted signals explained 22%–88% (60% on average) of the variance in the actual force, with the high gamma frequency band being the most informative feature.

On the other hand, kinetic information can also be recorded from deep brain areas using SEEG electrodes, and existing kinetic-based studies indicate the possibility of decoding continuous force using SEEG. For example, Murphy *et al* have shown that signals from deep cortical areas, including the central sulcus and the insular cortex, contain useful information to differentiate between the rest state and three different levels of hand grasping (Murphy *et al* 2016). Tan *et al* recorded nine participants with Parkinsonism performing a grasping task at different force amplitudes (Tan *et al* 2013). They observed that the subthalamic nucleus (STN) showed different ERS/ERD at different hand grasping levels. Specifically, they observed that beta and gamma modulations behaved differently as grasping force increased: as the force increased, beta suppression deepened, and then plateaued; however, gamma and high-frequency power (HFP) only increased. Furthermore, when force effort was categorized into discrete levels (1–10), power modulation in the beta band was the only independent predictor of force when effort levels <5, while an increase in gamma band activity was the only independent predictor when effort level $\geqslant 5$. In another study, Fischer *et al* recorded signals from the STN of eleven participants performing or imagining hand grasping for 5 s at different amplitudes (Fischer *et al* 2017). They found that both beta and gamma activities changed in accordance with the level of force throughout the grasping period.

Despite the kinetic information extracted from SEEG, the pioneering studies by Tan *et al* and Fischer *et al* did not evaluate sustained or prolonged force tasks, nor did they attempt to reconstruct the continuous force applied, both of which are critical requirements for BCIs in daily usage. Thus, to advance this line of research, the possibility of continuous grasp force decoding was investigated in the study using a prolonged grasping paradigm.

A deep learning model was proposed for force amplitude decoding in this paper. Five other methods were also tested to show the superior performance of the proposed method. Deep learning has already been demonstrated to be successful in BCI decoding tasks both for EEG and ECoG recordings. Various models have already been proposed for scalp EEG signal decoding tasks, such as TSception and EEGNet, both of which are capable of extracting information in temporal and spatial domains (Lawhern *et al* 2018, Ding 2020). For invasive recordings, Du *et al* showed that a combination of convolutional and recurrent neural networks achieved superior results in a finger trajectory regression task (Du *et al* 2018). However, neither of these models were designed specifically for SEEG, and they were not evaluated on SEEG signal data. In this work, a novel deep learning model was designed to decode the grasping force from SEEG signals.

Taken together, a continuous force control experiment was designed, which comprised two force increasing rates and two force amplitude targets over a duration of 15 s, resulting in four combinations (tasks): slow ascending light grasp; slow ascending hard grasp; fast ascending light grasp, and fast ascending hard grasp. SEEG signals were recorded during the experiment. Spectral responses for these four tasks were analyzed, then force amplitude was decoded using six methods, including a linear model, a partial least squares (PLS) model, an unscented Kalman Filter, and three deep learning models(two CNN models used in previous EEG studies and a CNN+RNN model proposed in this paper). By comparing the decoding result obtained by different methods, the novel deep learning architecture referred to as CNN+RNN in this paper, achieved the best decoding result.

2. Materials and methods

2.1. Participants and data recording

Five human participants with intractable epilepsy (referred to as *a*–*e*) were recruited in this study with written consent (in the case of two child participants, parental consent was granted). Participants were implanted with SEEG electrodes for presurgical assessment of seizure focus. All electrode parameters were determined solely by clinical need as part of the presurgical assessment. SEEG signals were acquired using a clinical recording system (EEG-1200C, Nihon Kohden, Irvine, CA) sampled at 2 kHz. The clinical profile of all participants is shown in table 1.

This study was reviewed and approved by the Ethical Committee of the University of Bath (Ethical Approval Reference No.: EP 20/21 050) and the Ethics Committee of Huashan Hospital (Shanghai, China) (Ethical Approval Reference No.: KY2019518).

2.2. Experimental protocol

A dynamic force control experiment was designed, in which the participants control their hand grasp force to follow a pre-designed curve presented with an LCD monitor. During the experiment, SEEG signals and force data were recorded simultaneously. The force data were sampled at 5 kHz using a force transducer (Forza, OT Bioelettronica) and the recorded force value was displayed to the participants in real-time using a moving forward black dot (*Y*-axis: force level, *X*-axis: time) on the screen. Before the experiment started, participants rested on a bed, keeping still and with their eyes fixed on the LCD monitor. Participants were given time to practice with the protocol, after which their MVC was measured. They were then asked to perform four tasks (i.e. 20% and 60% MVC in slow and fast ascending rates, making four combinations in total: 20% MVC slow; 20% MVC fast; 60% MVC slow; and 60% MVC fast, as shown in figure 1). There were four stages in each task: a 2 s preparation stage, a 3/9/1/3 s force increasing stage for tasks 1/2/3/4, a 2.5 s holding stage for all tasks, and a 7.5/1.5/9.5/7.5 s releasing stage for tasks 1/2/3/4 respectively. In the preparation stage, participants held the force transducer loosely without applying any force. During the force increasing and holding stage, participants were instructed to increase their grasp force following a moving black dot and held for 2.5 s (holding stage). Then the participants released the force transducer and relaxed for the remaining time (release stage). The entire experiment consisted of four sessions per participant, while each session contained 40 trials, where each of the four tasks was presented 10 times in random order. Overall, 160 trials were collected and the whole experiment lasted approximately 40 minutes for each participant.

2.3. Electrode localization

The participants had a total of 745 electrode contacts (rounded mean $\pm$ std: 149 $\pm$ 42 per participant) implanted. Each electrode shaft was 0.8 mm in diameter and contained 8–16 contacts (contact length was 2 mm) with 3.5 mm center-to-center spacing. In order to locate the electrodes, brain segmentation was first performed using pre-surgical MRI with Freesurfer (Fischl 2012). Then the 3D coordinates and the anatomical label of each electrode contact were obtained by co-registering post-surgical CT images with pre-surgical MRI using an open-source toolbox, iEEGview (Li *et al* 2019). The resulting 3D location of all electrode contacts is shown in figure 2.

2.4. Trial elimination

After data collection, each trial was evaluated by visual inspection, and trials with significant noise in force recording were rejected. Trials where the MSE (Mean Squared Error) between actual force and force in the experiment paradigm larger than 0.0084 (determined by visual inspection) were eliminated: this left 145, 149, 151, 152, and 150 trials for participants *a*, *b*, *c*, *d* and *e* respectively.

2.5. Signal preprocessing

Trials after elimination were further processed at this stage before decoding. First, both the raw SEEG and the force data were downsampled to 1 kHz. The SEEG signals were then bandpass filtered from 0.5 Hz to 150 Hz using a 4th order Butterworth filter before a Laplacian re-reference was applied (Li *et al* 2018). Subsequently, a notch filter was used to eliminate the 50 Hz line noise and its harmonics. The resulting SEEG data were used for further analyses, where the baseline was defined as 1 s long, starting 0.5 s after grip release for all trials.

2.6. Temporal-spectral representation

The temporal-spectral representation of the SEEG signals was obtained by performing time-frequency decomposition using the MNE toolbox (Gramfort

Table 1. Anonymized clinical profiles of participants in the study.

Participant ID	EZ	DH	EH	Gender	Age	RH	EL	NC	SR(Hz)
a	Right temporal lobe	R	L	F	23	RH	10	130	2000
b	Left temporal lobe	R	R	F	42	LH	10	137	2000
c	Left SMA	R	R	M	15	LH	8	110	2000
d	Right parietal lobe	R	L	M	15	RH	7	102	2000
e	Right motor area	R	L	F	35	RH	10	217	2000

Abbreviations for this table: EZ, epileptogenic zone; RH, recording hemisphere; SR, sampling rate; SMA, supplementary motor area; EL, number of electrode shafts; NC: number of contacts; DH, dominant hand; EH, experiment hand.

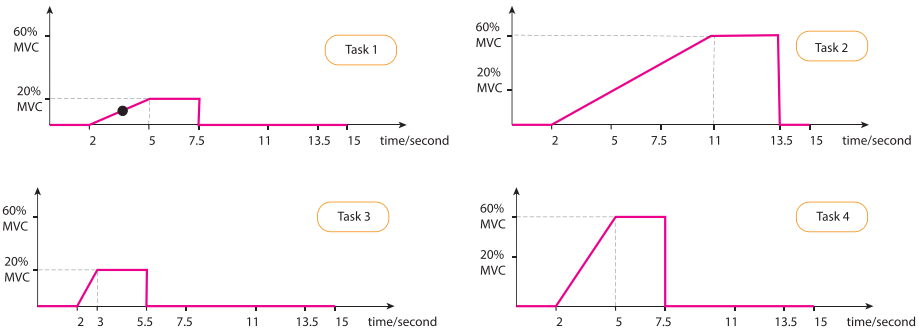

Figure 1. The experimental protocol. Participants began with a 2 s preparation stage for each task. For tasks 1 and 2, participants increased their grip force to reach the predefined force level (20% maximum voluntary contraction (MVC) and 60% MVC) following a moving-forward black dot for 3 and 9 s respectively, then held for 2.5 s before release. For tasks 3 and 4, participants increased their grip force to reach the predefined force level (20% MVC and 60% MVC) for 1 and 3 s respectively, then held for 2.5 s before release.

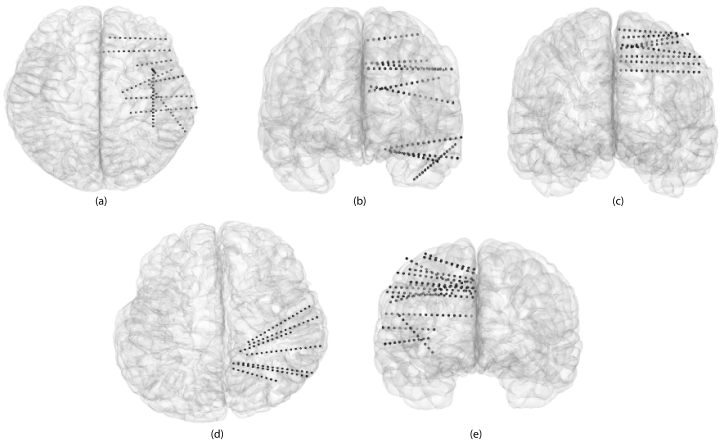

Figure 2. The 3D locations of the SEEG electrodes are shown for each individual brain model. Participants *a* and *d* are shown on the transverse plane, whereas participants *b*, *c* and *e* are shown on the coronal plane. The black dot indicates each SEEG contact.

et al 2013). In detail, FFT-based convolution using Morlet wavelets with wavelet cycles exactly equal to the frequency values were adopted to decompose data in the time-frequency domain ranging from 2 Hz to 150 Hz with 1 Hz frequency resolution. The power obtained was then normalized by first subtracting and then dividing by the mean power in the baseline, as illustrated in equation (1):

$$\Delta_{tf} = 100 \times \frac{activity_{tf} - \overline{baseline_f}}{\overline{baseline_f}}, \quad (1)$$

where $\overline{baseline_f}$ means the average power of frequency f. Δ_{tf} was then plotted to visualize the temporal-spectral evolution during the task.

2.7. ERS/ERD

ERD in low-frequency bands (LFBs) and ERS in high-frequency bands (HFBs) are well-known indicators of neural modulation in relation to the task. The possibility of discriminating tasks using ERS/ERD was investigated in this work. ERS/ERD was calculated in the same way as in Jiang *et al*

(2020). A detailed calculation step is presented below.

First, the signal ($v_c^i(t)$) from each channel c of trial i ($i \in (1, N)$, where N is the trial number), was band-pass filtered into a LFB (0.5–30 Hz) and a HFB (60–150 Hz) with a second-order Butterworth IIR zero-phase filter. Then a Hilbert transform was applied to the two bandpass-filtered signals, which were then squared to acquire the low-frequency power (LFP) and the HFP separately using equations (2) and (3):

$$LFP_c^i = (Hilbert(BP_{0.5-30\ \mathrm{Hz}}(v_c^i(t))))^2, \quad (2)$$

$$HFP_c^i = (Hilbert(BP_{60-150\ \mathrm{Hz}}(v_c^i(t))))^2, \quad (3)$$

where $t \in (0, 15\ \mathrm{s})$. The channel-wise power LFP_c and HFP_c were then calculated by averaging LFP_c^i and HFP_c^i across the N trials. Next, the power was normalized against the mean power of the baseline period to generate a scalar ERD and ERS using equations (4) and (5) respectively.

$$ERD_{LFP} = 10\log\left(\frac{\sum_{t\in T_{task}} LFB_c}{\sum_{t\in T_{ref}} LFB_c} \times \frac{N_{ref}}{N_{task}}\right), \quad (4)$$

$$ERS_{HFP} = 10\log\left(\frac{\sum_{t\in T_{task}} HFB_c}{\sum_{t\in T_{ref}} HFB_c} \times \frac{N_{ref}}{N_{task}}\right), \quad (5)$$

where T_{ref} and T_{task} are the baseline and task period. One ERS/ERD pair was obtained for each trial under each task. Then, the statistical significance was calculated between the ERS/ERD task pairs to evaluate if it is possible to separate tasks using ERS/ERD.

Statistical testing was performed using Wilcoxon rank-sum test without making assumptions of data distribution. The critical value was set to 0.05 in this paper.

2.8. Channel selection

The channel selection was performed before force decoding in order to reduce the total number of channels, thus reducing computational workload and avoiding over-fitting. The channel selection was done separately for each participant. First, all trials were split into training, evaluation, and testing datasets in a 70/10/20 manner for each participant. Then the top activated channels were selected using the activation index (*AI*) calculated from the training dataset (the same partitioning and channel were used in the decoding step). The *AI*s were calculated per channel using ERS/ERD according to equation (6):

$$AI_c = \frac{ERS_{HFP} - ERD_{LFP}}{ERS_{HFP}}. \quad (6)$$

AI_c is a value reflecting the modulation strength caused by grasping tasks that incorporates both high frequency and low-frequency information. Channels with a higher *AI* are more reactive to the task and were selected for subsequent frequency analysis and model training and prediction. The reduced channel number can also help to prevent overfitting, especially for methods that are not good at handling high-dimensional data. In this paper, channels with *AI* values >1.5 were chosen, leading to 9, 11, 14, 13, and 10 active contacts for participants *a*, *b*, *c*, *d*, and *e* respectively. This channel selection approach is in keeping with the previous hand gesture decoding study, which used the most active channels that were measured by the difference of frequency power between task and rest stages (Wang *et al* 2020).

2.9. Feature extraction

For both traditional and deep learning methods, frequency features in different bands were chosen because they have been proved to contain rich information about hand movement (Asher *et al* 2007, Kellis *et al* 2012, Bleichner *et al* 2016, Li *et al* 2018, Wang *et al* 2020). Further, high gamma and delta band (e.g. 0.5–4 Hz) ranges produced superior performance among the total ten frequency bands in a hand gesture decoding task using SEEG signals (Li *et al* 2022). Therefore, the power of five frequency bands in 0.5–4 Hz, 4–13 Hz, 13–30 Hz, 30–60, and 60–150 Hz were used as input features for both traditional and deep learning methods. In detail, each channel was band-pass filtered into five frequency bands using a sixth-order Butterworth filter, then power was computed by taking the squared value of the Hilbert transform of the filtered signal. The same features were extracted from the testing dataset to obtain the final prediction.

2.10. Grasp force reconstruction

In the last step, the extracted features were used to train six different decoding models, including three traditional approaches and three deep learning models.

Decoding and evaluation were performed separately for each decoding method and each participant. To ensure the equal distribution of different tasks among training, testing and validation datasets, `train_test_split` from `scikit-learn` was used to perform the partitioning (Pedregosa *et al* 2011). For the linear, PLS, and Kalman methods, models were trained on the training set and tested on the testing set. For deep learning models, evaluation datasets were used for hyperparameter tuning.

The MSE between the true force and the predicted results was used to evaluate the performance of each decoder. One MSE value was obtained for each trial using each method. In the end, MSEs for each decoder were obtained from testing trials, then the difference significance between decoders was tested using Wilcoxon rank-sum test as illustrated in section 2.7.

2.11. Traditional decoding methods

Three traditional decoding methods, including a linear model, a PLS, and an unscented Kalman filter were used in this study. The details of these methods are shown below.

2.11.1. Linear and PLS models

`LinearRegression` and `PLSRegression` from scikit-learn were used to implement the linear and PLS models (Pedregosa *et al* 2011). The model was trained on the training dataset before being tested on the testing dataset.

2.11.2. Unscented Kalman filter

An unscented Kalman filter was used to model the non-linearity relationship between force and neuronal signals by setting $N = 1$ for the Nth order unscented Kalman filter from Luu (2016). In this study, a quadratic equation (equation (7)) was used to represent the neural tuning model:

$$M_t = f_{t-l} + f'_{t-l} + \sqrt{f_{t-l}^2 + {f'_{t-l}}^2}, \tag{7}$$

where f_{t-l} and f'_{t-l} are the force amplitude and its first-order derivative, and $l = 100$ ms is the time delay. The remaining implementation of the unscented Kalman filter is the same as in the previous work (Luu 2016).

2.12. Deep learning methods

Three deep learning networks were implemented and compared in this study. The first two networks were a deep convolutional neural network (deepConv) and a shallow convolutional neural network (shallowConv) from Schirrmeister *et al* (2017), both of which have been proved to be comparable or superior (at the time of reporting) to state-of-the-art methods in EEG signal decoding tasks. The third method proposed in this paper is a novel CNN+RNN model, consisting of a temporal convolutional block, a spatial convolutional block, and a recurrent convolutional block. In the temporal and spatial convolutional blocks, multiple 1D convolutional layers were used to extract spectral-spatial features, inspired by TSception (Ding 2020) and EEGNet (Lawhern *et al* 2018). Then the extracted features were fed to a long-short-term memory (LSTM) RNN layer to output a scalar value.

The following sub-sections will explain the proposed novel CNN+RNN network in more detail.

2.12.1. Temporal convolutional block

The block consists of five 1D convolution layers. The length of the kernels is determined by the sampling rate of the SEEG data f_s. Here S_T^i, the kernel size of the i_{th} level ($i = 1, \ldots, 5$) from the temporal block in figure 3, can be defined as:

$$S_T^i = (1, \alpha^i \times f_s)(i = 1, \ldots, 5). \tag{8}$$

To capture frequencies at 2 Hz and higher, the first kernel length was set to be half of the sampling rate with $\alpha^1 = 0.5$, as in TSception and EEGNet. Similarly, α^2 was set to 0.25 to capture 4 Hz and above. For the five layers, α^i was set to (0.5, 0.25, 0.125, 0.0625, 0.03125). The input SEEG signal, X, for one batch is defined as:

$$X = (x^0, x^1, x^2 \ldots x^n), x^i \in \mathbb{R}^L, \tag{9}$$

where n is the total number of SEEG active channels and L represents the sampling points for one epoch trial and x^i is the i_{th} active channel. Then the output of each temporal convolution layer ($Z^i_{temoral}$) can be defined as:

$$Z^i_{temoral} = \text{AvgPool}(\text{ReLU}(\text{Conv1D}(X, S_T^i))), \tag{10}$$

where $i = [1, \ldots, 5]$ and S_T^i is the kernel size (equation (8)). Each temporal layer processes the 2-dimensional input into 3D data, then all five temporal outputs are stacked along the kernel dimension. This is further fed into the second component of the network (the spatial layer).

2.12.2. Spatial convolutional block

As with the temporal layer, the spatial layer has three 1D convolutional layers with different kernel sizes. Let the input to this layer be $X_T = X_{K \times C \times L}$, where K is the number of concatenated features or plans, C is the SEEG channels number, and L is the length of one trial of SEEG data. The output of this spatial layer Z^i_{cs} is then defined as:

$$Z^i_{spatial} = \text{AvgPool}(\text{ReLU}(\text{Conv1D}(X_{K \times C \times L}, S_S^i))), \tag{11}$$

where $i = [1, \ldots, 3]$ and S_S^i is the kernel size.

2.12.3. Recurrent convolutional block

After features were extracted from the previous layers, an LSTM RNN was used to aggregate the information from the window. Briefly, an LSTM updates a storage state through an input gate and a forget gate. An output gate then controls the information flowing to the next unit (Sak *et al* 2014). By allowing SEEG signal information to persist through time, the LSTM layer can aggregate historical information in the sliding window for the current prediction. ReLU (Agarap 2018) was used as the activation function in this model.

2.12.4. Training the deep learning models

The training was performed on the training dataset using back-propagation through time with MSE as loss metric. L1 regularization was applied to the parameters to avoid over-fitting (Tartaglione *et al* 2018). Together, the final loss function is defined as:

$$L(y, \widehat{y}) = \text{MSE}(y, \widehat{y}) + \lambda \sum | \theta_i |, \tag{12}$$

where y denotes the ground truth and $\widehat{y}$ as the prediction. θ_i represents the model parameters, and λ is

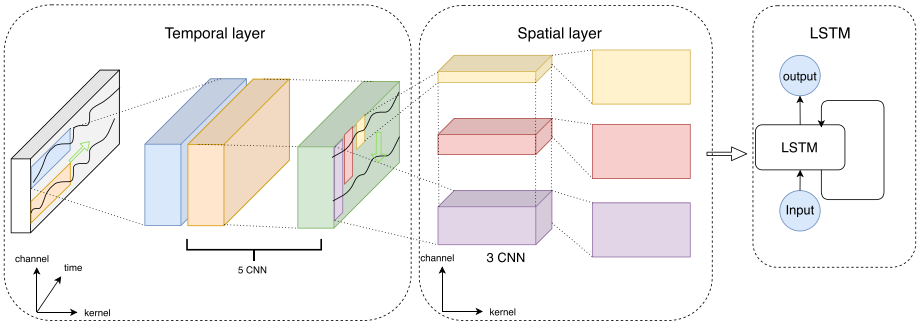

Figure 3. The proposed CNN+RNN architecture. The network consists of three blocks: a temporal convolution block; a spatial convolution block; and a recurrent convolution block (LSTM). A 2 s window with a 50 ms stride was used to slide through each trial and produce a prediction at each step. In the first stage, each 1D convolution layer produced a cubic output, which was then concatenated at the output layer. In the second stage, another set of 1D convolution layers was used to extract the spatial information. In the last stage, an LSTM was implemented to aggregate the extracted features in the window.

the regularization hyper-parameter which was set to 1.0×10^{-6} empirically.

During the training processes, a history 2 s sliding window with a 50 ms step length was used to generate a singular prediction value for each window at the current time point. Because history data was used for prediction, the decoding process was causal and therefore can be used for online testing. Then, the difference between true force and the prediction (loss) was used to update the network parameters.

3. Results

3.1. Temporal-spectral representation

An example of temporal-spectral representation, calculated as per section 2.6, of four tasks from participant *a* is presented in figure 4. As can be seen, there existed clear frequency modulations in different frequency ranges under different task stages. As shown in the upper half of each subplot, there were clear ERS and ERD in the high-frequency and low-frequency ranges. More specifically, there was a wider frequency range that showed ERS in the holding stage compared to that in the preparation and ascending stage. As shown in the lower half of each subplot, there was a wide separation between the mean ERS and ERD lines during the whole task stage, before ERS and ERD merged in the resting stage. This separation began around 1.5 s before force onset and extended for 0.5 s after force offset. This indicated that both high-frequency and low-frequency modulation recorded with the SEEG device can be sustained in a prolonged task. Visual inspection found no clear ERS/ERD in the resting stage.

Although a clear ERS/ERD can be seen in the example channel, visual inspection showed there was only 15%, 11%, 9%, 11%, and 21% of the total channels exhibiting strong ERS/ERD for participants *a*, *b*, *c*, *d*, and *e* respectively.

3.2. Significant difference in ERS/ERD

To investigate the discriminability of different tasks using their frequency response, the ERS/ERD of trials under two tasks were compared. ERS/ERD values were calculated as described in equations (4) and (5), and the statistical test procedure was performed using Wilcoxon rank-sum test as described in section 2.7. Bonferroni correction was applied to account for the multiple comparisons conducted by ERS and ERD.

An example result is presented in figure 5, using one example channel for each five participants. As indicated in the plot, it is possible to discriminate tasks using differences in frequency modulation. For example, the timings were the same for task 1 and task 4, while the maximum force level was higher for task 4. Accordingly, a larger absolute value of ERD was observed for task 4 compared to that of task 1, as shown in all five subplots. However, no consistent relationship was found between force level and ERS/ERD when considering all electrodes. This implies that it would be difficult (or that it may not be an effective strategy) to categorize force by simply mapping from ERS/ERD, which inspired us to use more powerful methods for force prediction, as outlined in the subsequent sections.

3.3. Grasp force reconstruction

To decode force using SEEG signals, 6 different models were trained on the training dataset using the methods described in sections 2.11 and 2.12, before predictions were obtained by feeding the testing dataset to the trained models. With the proposed CNN+RNN model, a concatenation of the predictions from 4 trials, taken from the 4 tasks for all five participants, is presented in figure 6. The results demonstrate that the prolonged force amplitude can be decoded from the SEEG recordings. Further, the predictions successfully fit the ascending, holding, and descending phases of the ground truth (the force that was actually applied); therefore, not only was force onset and offset accurately decoded but also the true force amplitude.

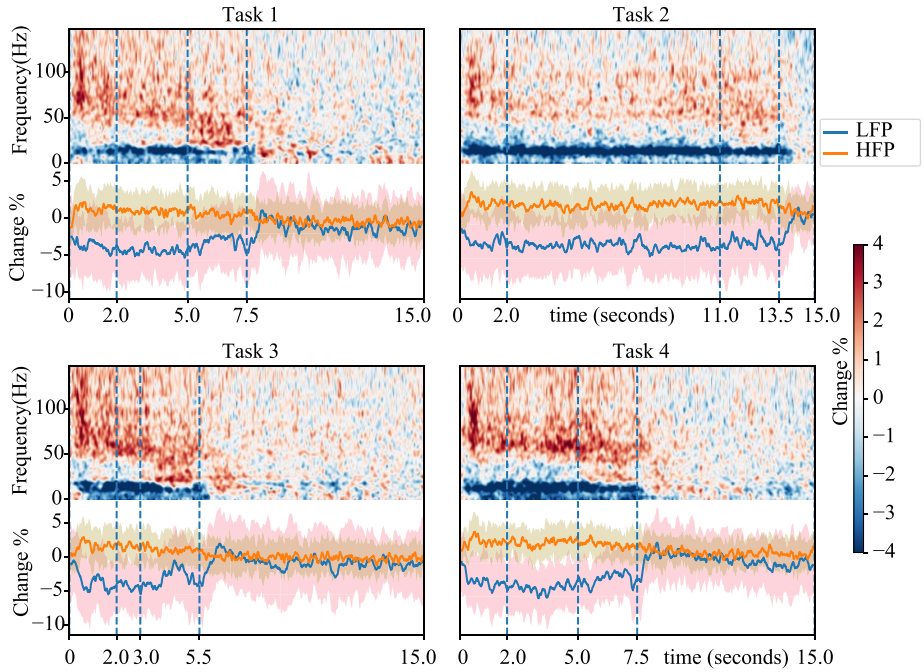

Figure 4. Temporal-spectral representations under four tasks recorded by one typical SEEG contact from participant *a*. The three vertical lines indicate the force onset, the start of the force holding stage, and the start of the release stage. The upper half of each subplot is the time-frequency decomposition, where the color indicates the percentage change from the baseline as described in section 2.6. The lower half of each subplot shows the time evolution for the average high-frequency power (HFP) (orange, 60–150 Hz) and low-frequency power (LFP) (blue, 0.5–30 Hz). The solid line and shaded area indicate the mean and standard deviation.

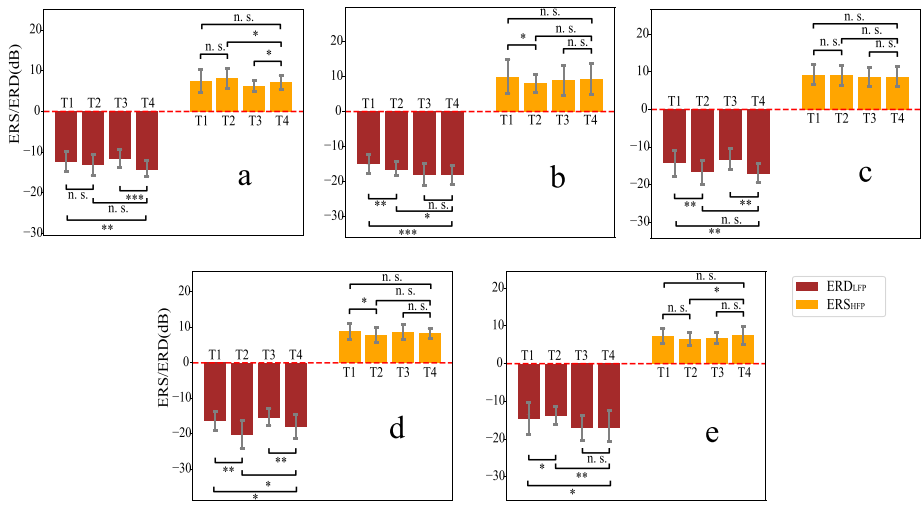

Figure 5. Significance analysis of ERS/ERD from one example active electrode of five participants (*a*–*e*). The vertical bars denote the mean ERD(red)/ERS(orange) value that averaged across all trials and the error bars denote the standard deviation. T1–T4 indicates the four different tasks. (***: $p < 0.0005$, **: $p < 0.005$, *: $p < 0.05$, n.s. denotes no significant difference. Wilcoxon rank-sum test.)

Predictions of all testing trials using all decoding methods can be found in supplementary figures 2–6 (available online at stacks.iop.org/JNE/19/026047/mmedia) in the appendix section. To evaluate the decoding performance, a chance level prediction MSE was calculated for each participant. The chance prediction was defined as the mean true force averaged across all trials of that participant. The MSEs of chance prediction are 0.5544, 0.5566, 0.5889, 0.6023 and 0.6021 for participants *a*, *b*, *c*, *d* and *e* respectively. For comparison: for participant *e*, the linear model, PLS, unscented Kalman, shallow network, deep network, and the proposed CNN+RNN achieved MSE losses of 0.15, 0.14, 0.19, 0.35, 0.30, and 0.05 respectively.

The mean decoding MSE error of six decoding methods under the four tasks averaged across participants is presented in figure 7(a). For the

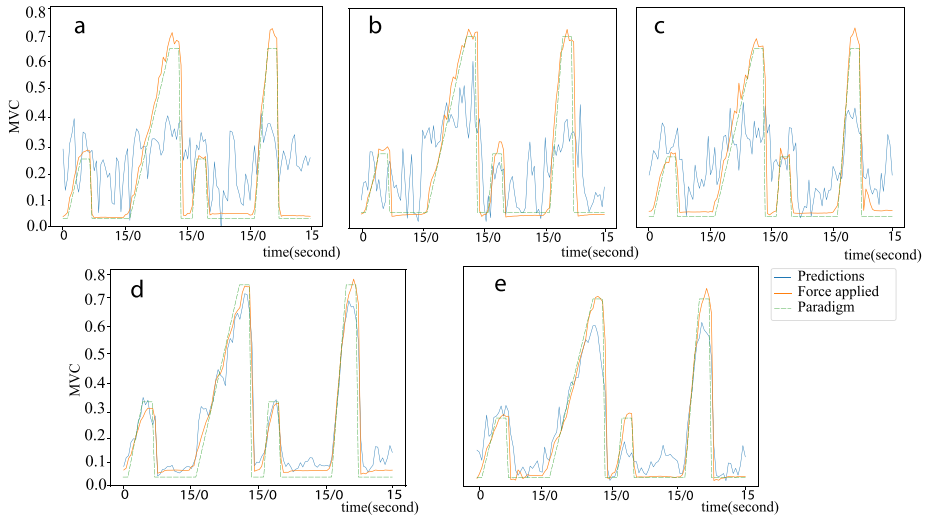

Figure 6. Force prediction for participants *a*–*e* using the CNN+RNN model. Each subplot was a concatenation of four trials predictions taken from 4 different tasks where each task lasts for 15 s. For the sake of visualization, only one trial per task is presented and concatenated for each participant. Complete predictions of all testing trials using all decoding methods can be found in the supplementary section.

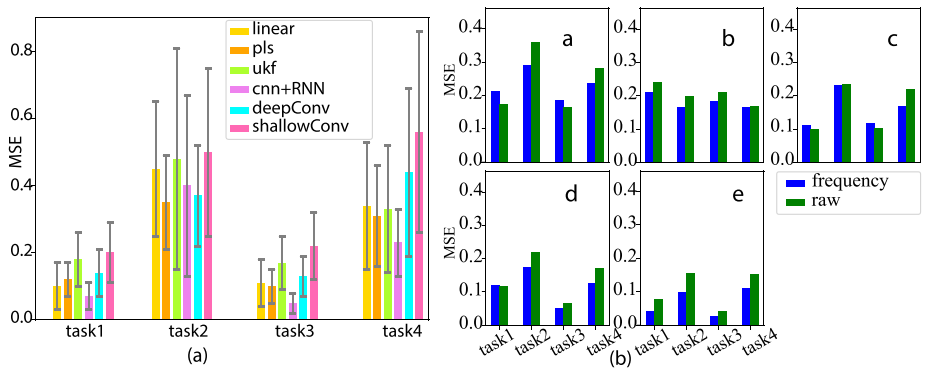

Figure 7. In subplot (a), the mean decoding MSE using six decoding methods is presented for each task averaged across participants. The vertical bars denote the mean MSE averaged across trials and participants. The top and bottom of each error bar are the 20th and 80th percentile, respectively. In subplot (b), the MSE decoding error using the CNN+RNN under the four tasks with two different inputs for all five participants (*a*–*e*) is presented. The bars show the mean MSE averaged across trials for each task. The blue bars represent MSE calculated using the extracted frequency features, while the green bars indicate the results calculated using the raw signals only.

MSEs calculated for each individual participant, please refer to supplementary figure 1. The results from the CNN+RNN network showed a lower MSE compared with all other models tested under all four tasks. To statistically test the difference, a Wilcoxon rank-sum test was conducted. The significance p-values between the linear/PLS/UKF/deepConv/shallowConv and CNN+RNN models are $1.03 \times 10^{-03}/1.04 \times 10^{-04}/2.27 \times 10^{-03}/4.97 \times 10^{-02}/2.76 \times 10^{-03}$ respectively.

Additionally, to evaluate the ability of the proposed deep learning model to learn from raw SEEG signals, similar to the input of EEGNet (Lawhern *et al* 2018) and many other EEG-based deep learning studies (Schirrmeister *et al* 2017, Mousavi 2019, Wu *et al* 2019, Rashid *et al* 2020), raw data was used to retrain and test the CNN+RNN network. The comparison between MSEs using different inputs for all five participants was presented in figure 7(b), which indicated a comparable performance using different inputs, demonstrating the feasibility of using deep learning methods on raw SEEG recordings.

4. Discussion

This work has investigated the possibility of decoding continuous grasping force using SEEG signals for the first time by employing a novel paradigm with two different increasing slopes and two force levels. Additionally, the temporal-spectral representation of neural responses was investigated under different force control tasks. Furthermore, after evaluation of the decoding performances using six different decoders, a novel CNN+RNN architecture was proposed in this work which produced the best decoding performance for continuous force prediction.

4.1. ERS/ERD

The temporal-spectral representation indicated that there was a clear ERS/ERD throughout the task stage which started around 2 s before force onset and 1 s after force offset. Additionally, there was a persistent ERS/ERD during the holding stage as shown in figure 4 of an example electrode. However, different behavior of ERS/ERD was observed from other electrodes. For example, 12 out of all 102 electrodes from subject *d* showed clear and sustained (not reduced) ERD during the holding stage. For the other electrodes which exhibited ERD, the strength was reduced but never returned to baseline level. These 12 electrodes were distributed in different areas, including the parietal and cingulate cortex. On the other hand, 3 out of 102 electrodes showed clear and sustained ERS during the holding stage similar to ERS in figure 4. These three electrodes were all located in the supramarginal gyrus. The similar sustained ERD can be found in all subjects, while sustained ERS was only found in subjects *d*. This behavior is different from two previous ECoG-based studies which showed that power modulation failed to sustain while the participants held their grasping force (Branco *et al* 2019, Jiang *et al* 2020). Jiang *et al* used a hand grasp paradigm containing a force holding stage of 2–3 s and found that ERS/ERD returned to baseline immediately after force onset in a force holding paradigm (Jiang *et al* 2020). They reported that the fluctuations of ERS in the HFB primarily, and of ERD in the LFB to a lesser extent, correlated with the time-course of the first time-derivative of force (yank), rather than with force itself. The same observation was found in another ECoG study where a random 1–2 s holding stage was utilized in a hand grasping paradigm (Branco *et al* 2019). They found that the HFB power failed to exhibit a sustained response related to a constant force, and a similar conclusion was reached regarding the relationship between the force yank and the HFB power. Another study using scalp EEG also demonstrated that frequency modulation failed to sustain when subjects held their hand gesture (Cassim *et al* 2000). The different ERS/ERD behavior demonstrated the advantage of SEEG in recording signals from deep brain regions to reveal different response in different locations. However, due to the limited data in this paper, further study is needed to understand this difference.

4.2. Channel selection

In this study, the active channels were chosen according to their frequency response evaluated by the activation index (*AI*). The same channel selection procedure was used in other SEEG studies (Li *et al* 2017). Although a promising decoding result can be obtained, it does not mean that the active channels selected by the frequency response are optimal. Two possible situations undermine this method. First, the features were extracted and averaged from several fixed frequency bands, which might overlook the real informative feature in narrow bands. Second, activated channels are not necessarily helpful for the decoding task. For example, a channel might be activated in all task stages but show no discrimination between high and low forces. Thus a future study that investigates the optimal channel selection method in SEEG-based BCIs would be beneficial.

4.3. Force prediction

To further test whether recordings made with SEEG can be used for continuously changing force prediction, or can only distinguish discrete low/high force levels as demonstrated in other SEEG studies, this study designed a paradigm where the force changed within a single task and between different tasks. For tasks 1 and 4, the same experimental timings were followed across all stages except for the target forces; regardless, the decoder is still able to successfully predict two different force levels. This shows that the CNN+RNN model is decoding the real continuous force rather than a binary 'low force' or 'high force' condition. On the other hand, decoded force using other methods showed rhythmic oscillations and failed to reflect the force targets under different tasks. Therefore, for these methods, it is the force states (rest vs. grasping) that have been decoded instead of the real force.

4.4. Varied decoding performance

Whilst grasp force can be decoded from SEEG signals, the results varied amongst the participants and amongst different tasks for the same participants as shown in figures 6 and 7(a). As the electrodes were placed based on clinical needs, the performance difference may arise from the fact that the electrodes were placed in different locations. For participant *e*, the superior decoding performance might be because that part of the selected electrode was located in the left motor area which is a central area for movement control. Grasp force can also be decoded partially from participant *d*, possibly because two electrodes were selected from the posterior parietal cortex (PPC) which is part of the motor system and poses a fundamental role in visuomotor transformations (Fogassi and Luppino 2005). Wang *et al* also proved that signals from the PPC are helpful for decoding in a hand gesture classification task (Wang *et al* 2020). On the other hand, the average MSEs under tasks 2 and 4 were higher than that of tasks 1 and 3, as presented in figure 7(a). The worse performance (higher MSE) may be due to the higher force target in tasks 2 and 4. When taking into consideration another SEEG-based force study (Tan *et al* 2013) which showed that beta suppression deepened and then plateaued as the force increased, both studies suggest difficulties in decoding a high force, raising the question of whether there exists an upper limit for which force can be decoded using the SEEG recordings.

4.5. Deep learning model

When using deep learning neural networks, previous works have used raw data as the input (e.g. without preprocessing or filtering) and postulated that the convolution kernel will perform the feature extraction automatically by simulating spectral and spatial filters. For example, in the TSception network, two components—a spectral and a spatial learner—were used to automatically learn the spectral and spatial filters simultaneously (Ding 2020). In order to improve the overall performance, the authors designed convolution kernels with different lengths to simulate different filters. To investigate whether such designs with different kernel lengths can be used on SEEG data, this work also used raw data as input (alongside retraining the network). It showed that the decoding accuracy when using raw data as the model input was comparable to accuracy when using frequency band features, as shown in figure 7(b). This result demonstrated the ability of the deep learning model to learn from raw SEEG recordings.

5. Conclusion

In this article, a grasping force control paradigm that comprises two force targets and two force ascending rates was designed to investigate the possibility of continuous force decoding using SEEG signals. Temporal-spectral representation was first analyzed which showed very different spectral modulation in sustained grasping tasks compared with previous ECoG or EEG studies. Next, using six decoding methods, we demonstrated that prolonged grasping force can be decoded with high accuracy and a CNN+RNN deep learning method achieved the best decoding accuracy. The decoded force reflected the true 'rest or task' status, as well as the continuously changing amplitude under different ascending rates and force targets. The result presented in this work will help to better support the usage of SEEG signals for BCI applications.

Data availability statement

The data generated and/or analysed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request.

Acknowledgments

This work was supported by grants from the National Key Research and Development Program of China (No. 2018YFB1307301), the National Natural Science Foundation of China (No. 91848112), the China Postdoctoral Science Foundation (Grant No. 20Z102060158), and the Medical and Engineering Cross Foundation of Shanghai Jiao Tong University (Grant No. AH0200003).

ORCID iDs

Xiaolong Wu https://orcid.org/0000-0003-3943-4183
Guangye Li https://orcid.org/0000-0003-2530-3916
Scott Wellington https://orcid.org/0000-0002-8384-8474
Liang Chen https://orcid.org/0000-0002-9102-8193

References

Agarap A F 2018 Deep learning using rectified linear units (ReLU) (arXiv:1803.08375)
Asher I, Stark E, Abeles M and Prut Y 2007 Comparison of direction and object selectivity of local field potentials and single units in macaque posterior parietal cortex during prehension *J. Neurophysiol.* **97** 3684–95
Bleichner M G, Freudenburg Z V, Jansma J M, Aarnoutse E J, Vansteensel M J and Ramsey N F 2016 Give me a sign: decoding four complex hand gestures based on high-density ECoG *Brain Struct. Funct.* **221** 203–16
Branco M P, Geukes S H, Aarnoutse E J, Vansteensel M J, Freudenburg V Z and Ramsey N F 2019 High-frequency band temporal dynamics in response to a grasp force task *J. Neural Eng.* **16** 056009
Cassim F, Szurhaj W, Sediri H, Devos D, Bourriez J, Poirot I, Derambure P, Defebvre L and Guieu J 2000 Brief and sustained movements: differences in event-related (de)synchronization (ERD/ERS) patterns *Clin. Neurophysiol.* **111** 2032–9
Chen C, Shin D, Watanabe H, Nakanishi Y, Kambara H, Yoshimura N, Nambu A, Isa T, Nishimura Y, Koike Y 2013 Prediction of hand trajectory from electrocorticography signals in primary motor cortex *PLoS One* **8** 12
Ding Y 2020 TSception: a deep learning framework for emotion detection using EEG (arXiv:2004.02965v2)
Du A, Yang S, Liu W and Huang H 2018 Decoding ECoG signal with deep learning model based on LSTM *TENCON 2018—2018 IEEE Region 10 Conf.* pp 0430–5
Farrokhi B and Erfanian A 2020 A state-based probabilistic method for decoding hand position during movement from ECoG signals in non-human primate *J. Neural Eng.* **17** 026042
Fischer P *et al* 2017 Subthalamic nucleus beta and gamma activity is modulated depending on the level of imagined grip force *Exp. Neurol.* **293** 53–61
Fischl B 2012 FreeSurfer *NeuroImage* **62** 774–81
Flint R D *et al* 2014 Extracting kinetic information from human motor cortical signals *NeuroImage* **101** 695–703
Flint R D, Rosenow J M, Tate M C and Slutzky M W 2017 Continuous decoding of human grasp kinematics using epidural and subdural signals *J. Neural Eng.* **14** 016005
Fogassi L and Luppino G 2005 Motor functions of the parietal lobe *Curr. Opin. Neurobiol.* **15** 626–31
Gramfort A *et al* 2013 MEG and EEG data analysis with MNE-python *Front. Neurosci.* **7** 267
Herff C, Krusienski D J and Kubben P 2020 The potential of stereotactic-EEG for brain-computer interfaces: current progress and future directions *Front. Neurosci.* **14** 123
Huang W, Zhang P, Yu T, Gu Z, Guo Q and Li Y 2021 A p300-based BCI system using stereoelectroencephalography and its application in a brain mechanistic study *IEEE Trans. Biomed. Eng.* **68** 2509–19
Jiang T, Pellizzer G, Asman P, Bastos D, Bhavsar S, Tummala S, Prabhu S and Ince N F 2020 Power modulations of ECoG alpha/beta and gamma bands correlate with time-derivative of force during hand grasp *Front. Neurosci.* **14** 100

Kellis S, Hanrahan S, Davis T, House P A, Brown R and Greger B 2012 Decoding hand trajectories from micro-electrocorticography in human patients *2019 41st Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)* pp 4091–4

Lawhern V J, Solon A J, Waytowich N R, Gordon S M, Hung C P and Lance B J 2018 EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces *J. Neural Eng.* **15** 056013

Li G *et al* 2022 Assessing differential representation of hand movements in multiple domains using stereo-electroencephalographic recordings *NeuroImage* **250** 118969

Li G, Jiang S, Chen C, Brunner P, Wu Z, Schalk G, Chen L and Zhang D 2019 iEEGview: an open-source multifunction GUI-based Matlab toolbox for localization and visualization of human intracranial electrodes *J. Neural Eng.* **17** 016016

Li G, Jiang S, Paraskevopoulou S E, Wang M, Xu Y, Wu Z, Chen L, Zhang D and Schalk G 2018 Optimal referencing for stereo-electroencephalographic (SEEG) recordings *NeuroImage* **183** 327–35

Li G, Jiang S, Xu Y, Wu Z, Chen L and Zhang D 2017 A preliminary study towards prosthetic hand control using human stereo-electroencephalography (SEEG) signals *2017 8th Int. IEEE/EMBS Conf. on Neural Engineering (NER)* pp 375–8

Luu T P 2016 Unscented Kalman filter for neural decoding of human treadmill walking from non-invasive electroencephalography *2016 38th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)* pp 1548–51

Meng K, Grayden D B, Cook M J, Vogrin S and Goodarzy F 2021 Identification of discriminative features for decoding overt and imagined speech using stereotactic electroencephalography *2021 9th IEEE Int. Winter Conf. on Brain-Computer Interface (BCI)* pp 105–10

Mousavi S 2019 SleepEEGNet: automated sleep stage scoring with sequence to sequence deep learning approach *PLoS One* **14** e0216456

Murphy B A, Miller J P, Gunalan K and Ajiboye A B 2016 Contributions of subsurface cortical modulations to discrimination of executed and imagined grasp forces through stereoelectroencephalography *PLoS One* **11** e0150359

Nakanishi Y, Yanagisawa T, Shin D, Fukuma R, Chen C, Kambara H, Yoshimura N, Hirata M, Yoshimine T, Koike Y 2013 Prediction of three-dimensional arm trajectories based on ECoG signals recorded from human sensorimotor cortex *PLoS One* **8** e72085

Orsborn A L, Wang C, Chiang K, Maharbiz M M, Viventi J and Pesaran B 2015 Semi-chronic chamber system for simultaneous subdural electrocorticography, local field potentials and spike recordings *2017 8th Int. IEEE/EMBS Conf. on Neural Engineering (NER)* pp 398–401

Pedregosa F *et al* 2011 Scikit-learn: machine learning in Python *J. Mach. Learn. Res.* **12** 2825–30

Rashid M, Islam M, Sulaiman N, Bari B S, Saha R K and Hasan M J 2020 Electrocorticography based motor imagery movements classification using long short-term memory (LSTM) based on deep learning approach *SN Appl. Sci.* **2** 1–7

Sak H, Senior A and Beaufays F 2014 Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition (arXiv:1402.1128)

Sburlea A I and Müller-Putz G R 2018 Exploring representations of human grasping in neural, muscle and kinematic signals *Sci. Rep.* **8** 16669

Schalk G and Leuthardt E C 2011 Brain-computer interfaces using electrocorticographic signals *IEEE Rev. Biomed. Eng.* **4** 140–54

Schirrmeister R T, Springenberg J T, Fiederer L D J, Glasstetter M, Eggensperger K, Tangermann M, Hutter F, Burgard W and Ball T 2017 Deep learning with convolutional neural networks for EEG decoding and visualization *Hum. Brain Mapp.* **38** 5391–420

Schwarz A, Ofner P, Pereira J, Sburlea A I and Müller-Putz G R 2017 Decoding natural reach-and-grasp actions from human EEG *J. Neural Eng.* **15** 016005

Shin D, Watanabe H, Kambara H, Nambu A, Isa T, Nishimura Y and Koike Y 2012 Prediction of muscle activities from electrocorticograms in primary motor cortex of primates *PLoS One* **7** e47992

Tan H *et al* 2013 Complementary roles of different oscillatory activities in the subthalamic nucleus in coding motor effort in Parkinsonism *Exp. Neurol.* **248** 187–95

Tartaglione E, Lepsøy S, Fiandrotti A and Francini G 2018 Learning sparse neural networks via sensitivity-driven regularization (arXiv:1810.11764)

Wang M, Li G, Jiang S, Wei Z, Hu J, Chen L and Zhang D 2020 Enhancing gesture decoding performance using signals from posterior parietal cortex: a stereo-electroencephalograhy (SEEG) study *J. Neural Eng.* **17** 046043

Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain-computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91

Wu H, Niu Y, Li F, Li Y, Fu B, Shi G and Dong M 2019 A parallel multiscale filter bank convolutional neural networks for motor imagery EEG classification *Front. Neurosci.* **13** 1275