10.3217/978-3-99161-093-9-037

Movement-Onset-Aligned EEG for Generalizable Arm and Hand Movement Decoding: A Large-Scale Multi-Experiment Dataset

K. Sterk[1], K. Kostoglou[1], G.R. Müller-Putz[1,2]

[1]Institute of Neural Engineering, Graz University of Technology, Graz, Austria
[2]BioTechMed Graz, Graz, Austria

E-mail: gernot.mueller@tugraz.at

ABSTRACT: This work presents a curated, movement-onset-aligned electroencephalography (EEG) dataset comprising nine experiments with 188 able-bodied participants performing hand and arm movement execution tasks as well as rest conditions. Movement onsets were aligned using auxiliary sensors. All recordings were uniformly processed using causal filtering to ensure compatibility with online brain-computer interface (BCI) applications. Further processing steps include resampling, independent component analysis (ICA) for removal of ocular and noise components, and strict epoch rejection. Across experiments, the datasets include recordings with different channel configurations, covering 94 distinct EEG channels. In total, the dataset comprises over 95,000 annotated trials with detailed movement labels and is provided in a standardized HDF5 structure to facilitate flexible data selection in machine learning workflows. Signal analyses confirm the presence of physiologically plausible movement-related patterns over sensorimotor regions. This resource aims to support the development of cross-participant and cross-experiment generalizing movement decoders and transfer learning in EEG-based BCIs. The data will be released in the near future.

INTRODUCTION

Electroencephalography (EEG)-based brain-computer interfaces (BCIs) provide a direct communication pathway between the brain and external devices by translating neural activity to control signals. In motor-related BCIs, control signals are commonly derived from neural processes associated with voluntary motor activity, including movement execution [1], attempted movement [2], and motor imagery [3]. Such systems have attracted significant attention as assistive technologies for individuals with severe motor impairments. For example, EEG-based BCIs for persons with impairments of the upper limbs can be used to control (neuro)prosthetic devices [4].

In recent years, deep learning approaches have increasingly been explored for EEG decoding tasks due to their ability to learn hierarchical representations from training data. They have demonstrated promising results in a variety of BCI applications [5] by modeling nonlinear relationships and spatial-temporal dependencies within the data. However, deep learning models typically require large amounts of training data to achieve robust performance and to generalize well. In practical BCI applications, training data are usually collected during calibration sessions with the intended user. These sessions are often limited in duration, resulting in relatively small subject-specific datasets, especially for motor-impaired users where prolonged calibration sessions can be physically and mentally demanding. Consequently, insufficient training data remain a major challenge for developing reliable decoding models for EEG-based BCIs.

One promising strategy to address this challenge is the usage of generic deep learning classifiers pre-trained on a larger population [6]. Such models may capture generalizable neural patterns associated with motor processes and can subsequently be adapted to individual users through transfer learning and fine-tuning with a relatively small amount of subject-specific data. Despite the growing interest in data-driven methods for EEG decoding, the availability of large, well-curated datasets suitable for developing and evaluating such approaches remains limited. In particular, datasets that integrate recordings across multiple experimental paradigms and participants while maintaining consistent preprocessing and annotation are still relatively scarce. To support research in this direction, the present work introduces a curated large-scale EEG dataset intended to facilitate the development of machine learning and specifically deep learning methods for motor-related EEG decoding in BCI applications.

MATERIALS AND METHODS

Raw Data: Data from nine experiments were used for the creation of the dataset. The experiments were conducted at the Institute of Neural Engineering, Graz University of Technology, between 2017 and 2025. The individual raw datasets were assigned identifiers A to I. All experiments comprise EEG recordings acquired during right-hand and/or arm movement execution tasks performed by able-bodied participants. The signals were acquired using gel-based EEG systems. While some experiments focused exclusively on either arm or hand movements, others used combined motions. Movements were executed in response to visual cues. Most experiments additionally included resting state recordings, either active rest or passive rest. Active rest denotes trial-based

 10.3217/978-3-99161-093-9-037

rest periods explicitly cued to remain at rest, whereas passive rest refers to continuous resting periods without trial-specific cues. The original recordings differ in the number of EEG electrodes and the sampling frequency. An overview of all experiments is provided in Table 1.

Table 1: Overview of the datasets used with information on their original publication, number of participants, number of used EEG channels, and types of movements included

ID	Source	# Participants	# EEG Channels	Movement Types
A	[1]	15	61	Arm, Hand, Rest
B	[7]	15	58	Arm&Hand, Rest
C	[8]	20	60	Arm, Rest
D	[9]	20	60	Arm, Rest
E	[10]	15	61	Arm&Hand, Rest
F	[11]	34	61	Arm&Hand
G	[12]	25	60	Hand, Rest
H	[13]	22	60	Hand, Rest
I	[14]	22	60	Hand, Rest

Movement Onset Detection: Due to differences in the experimental paradigms, the time interval between cue onset and actual movement execution varied substantially across experiments. To ensure consistent alignment of the trials to the true movement onset, precise onset timings were extracted from auxiliary movement tracking systems. These systems include exoskeletons [1], data gloves [1][10], camera-based motion capture systems [8][9][12][13][14], pressure buttons [10], and Myo armbands [11]. Movement onsets were determined using threshold-based detection applied to task-relevant metrics, such as hand velocity or acceleration. Trials in which no movement could be reliably detected were excluded from further analysis. For the active rest condition, a virtual movement onset was defined to account for the temporal influence of the visual cue. To this end, the average delay d_{MO} between cue onset (CO) and movement onset (MO) was calculated for each participant during movement trials. The virtual onset for rest trials was then defined as $t_{MO,rest} = t_{CO,rest} + d_{MO} + r \cdot std(d_{MO})$, where $r \in [-1,1]$ is a uniformly sampled random factor and $std(d_{MO})$ denotes the standard deviation of movement onset delays. The random delay was introduced to mimic the variability in participants' reaction times observed during movement trials.

EEG Preprocessing: The EEG data preprocessing pipeline was designed with two main objectives: (i) to transform data from heterogeneous experiments into a uniform structure, and (ii) to remove strong artifacts and noise while preserving task-relevant EEG information. Processing was performed using MNE-Python [15]. All filters were chosen as causal filters to allow the dataset to be used for online BCI applications. First, line noise was attenuated using notch filters at 50 Hz and 100 Hz. Subsequently, a 0.5 Hz to 90 Hz bandpass filter was applied. The filters were implemented as 4^{th}-order causal Butterworth filters. EEG channels identified as floating (i.e., exhibiting strongly reduced band power across frequencies) were interpolated using spherical spline interpolation. The data were then resampled to 256 Hz using a causal anti-aliasing finite impulse response lowpass filter with Kaiser window [16]. Afterwards, independent component analysis (ICA) with Picard algorithm [17] was computed to identify and remove ocular artifacts. Further noise components, such as muscle artifacts and channel noise, were addressed using the MNE_ICALabel extension [18] for automated component classification. Components with a predicted artifact probability greater than 95% were removed. Additionally, components with class probability to represent noise or artifacts greater than 85% were removed if confirmed by visual inspection. Next, the data were segmented into 8-s epochs around the movement onsets, comprising 3 s pre-onset and 5 s post-onset. An additional epoch rejection step was performed to exclude residual noisy epochs. For each participant, the standard deviation was computed per channel across the entire recording. An epoch was rejected if any channel exhibited (i) a standard deviation exceeding five times or (ii) an absolute amplitude exceeding eight times the channel's global standard deviation. Finally, epochs were mapped to a uniform EEG channel layout comprising all electrodes used across experiments. Channels not present in the respective experiment were inserted and filled with NaN values to preserve structural consistency.

Dataset Structure: The final curated dataset is structured as hierarchical data format (HDF), version 5, files. To limit file size, each experiment is stored in a separate HDF5 file. Each file contains the processed and epoched EEG data, information on the timings of the movement onsets, and additional metadata stored as group or file attributes. Each experiment file *experiment_X.hdf5* is organized hierarchically as follows:

- File-level attributes
  - "Bandpass Filter Frequencies (Hz)" [Array of float]
  - "Notch Filter Frequencies (Hz)" [Array of integer]
  - "Channel Names (standard_1005)" [Array of string]
  - "GND" [string]
  - "Reference" [string]
  - "Sampling Frequency (Hz)" [integer]
  - "Event Codes" [Array of integer]
  - "Event Codes Meanings" [Array of string]
  - "Experiment ID" [string]
- Participant Group *Participant_X00*
- Session Group *Session_0* (only present in dataset I)
- Movement Group, either *Hand Movement*, *Arm Movement*, *Arm and Hand Movement* or *Rest*
  - "Movement Group" [string]
  - "Movement Details" [Array of string]
  - "Participant ID" [string]

Within a movement group, two data objects are stored:

- "eeg_data" [Array of float]: Epoched EEG data, of shape $(n_{epochs} \times n_{channels} \times n_{timepoints})$
- "events" [Array of integer]: Indicating timings of movement onsets, of shape $(n_{epochs} \times n_{timepoints})$

For dataset I, data are subdivided into sessions for each participant, reflecting the original experimental design, where multiple sessions were recorded per participant. The primary data objects are "eeg_data" and "events",

10.3217/978-3-99161-093-9-037

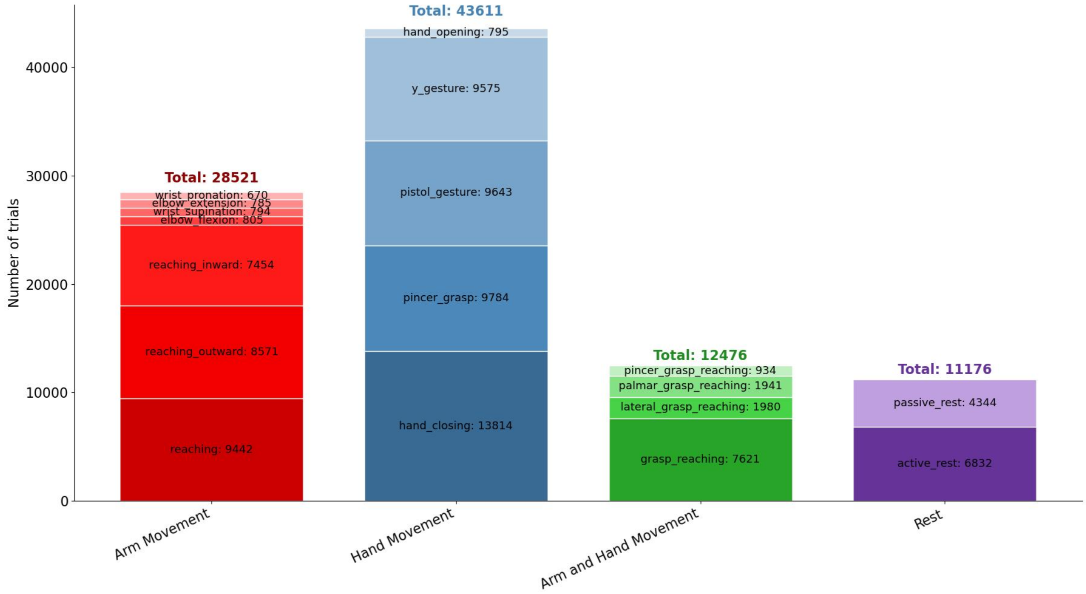

Figure 1: Number of trials per movement group of the processed dataset with occurrence of the detailed movement descriptions.

with "eeg_data" containing the processed EEG epochs and "events" array marking movement onsets (coded as 1 for movement onset, 0 otherwise). Note that although all epochs are centered around movement onsets and span 3 s pre-onset and 5 s post-onset, an epoch may contain multiple movement onset events. This occurs when inter-trial intervals are smaller than the chosen epoch size of 8 s.

RESULTS

Dataset Size: The dataset includes data from 188 participants and comprises 10 HDF5 files from 1.18 GB to 6.97 GB in size, resulting in a total volume of 38.5 GB.

Movement Groups: The EEG trials are organized into four movement groups with detailed movement labels:

- **Hand Movement:** hand closing, hand opening, pincer grasp, pistol gesture, y gesture
- **Arm Movement**: elbow extension, elbow flexion, reaching, reaching inward, reaching outward, wrist pronation, wrist supination
- **Arm and Hand Movement**: grasp reaching, lateral grasp reaching, palmar grasp reaching, pincer grasp reaching
- **Rest**: active rest, passive rest

Table 2: Number of trials per movement group of the processed dataset.

Movement Group	Number of Trials
Arm Movement	28521
Hand Movement	43611
Arm and Hand Movement	12476
Rest	11176
Total	**95784**

Trials in the Arm and Hand Movement group represent combinations of hand and arm actions (e.g., reaching combined with grasping). When a further differentiation of grasping or reaching movements was not possible, trials were labeled generically as grasp or reaching. During preprocessing, 22.7% of trials were excluded, resulting in a total of 84,608 movement trials and 11176 rest trials. An overview of the number of trials per movement group is provided in Table 2, while the distribution of detailed movement labels is shown in Figure 1.

EEG Channels: Across experiments, the datasets include recordings with different EEG channel configurations (58-64 channels), covering 94 distinct channels overall, following the extended international 10-05 system [19]. A subset of 31 channels is present in all trials. The availability of individual channels relative to the total number of trials is illustrated in Figure 2.

Data Quality: The effect of data preprocessing

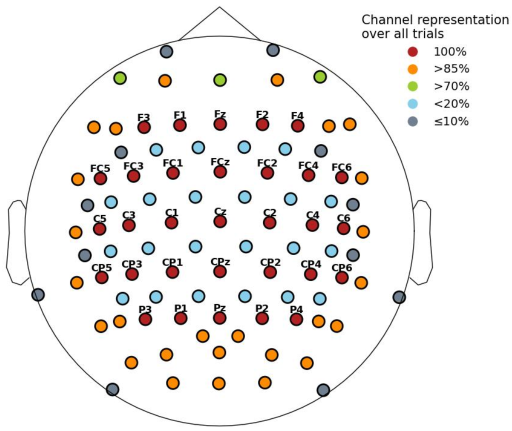

Figure 2: Channel representation with respect to the overall number of trials in the dataset. The channels in red are present in all trials. No channels had availability between 20% and 70%.

10.3217/978-3-99161-093-9-037

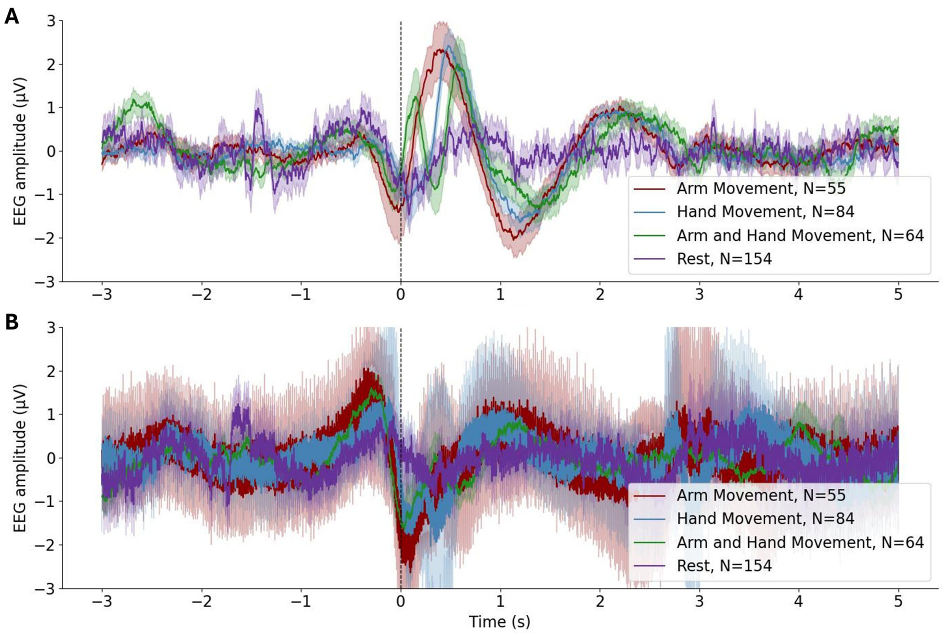

Figure 3: ERP of channel C1 per movement group for the fully and minimally processed data averaged over all trials of all participants with 95% confidence interval. N indicates the number of participants per movement group. Time point 0 s marks the movement onset. **(A)** ERP of the fully processed dataset. **(B)** ERP of the minimally processed data.

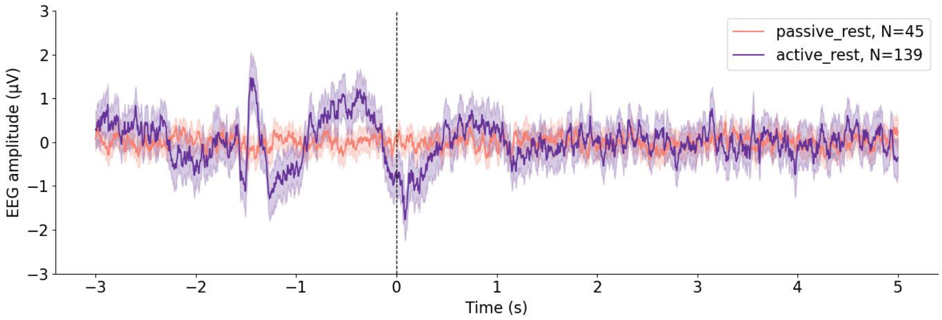

Figure 4: Activity of channel C1 of the fully processed data for movement details active and passive rest for the processed data averaged over all trials of all participants with 95% confidence interval. N indicates the number of participants per rest group. Time point 0 s marks the virtual rest onset.

was evaluated by comparing the event-related potentials (ERPs) derived from the fully processed dataset to those obtained from minimally processed data high-pass filtered at 0.5 Hz using a non-causal 4^{th}-order Butterworth filter, without further artifact correction or trial rejection. For the minimally processed dataset a non-causal filter was used for the purpose of offline analysis and to preserve the temporal morphology of the ERPs by avoiding phase distortions or latency shifts. In Figure 3, the ERPs of channel C1 per movement group for the fully processed dataset (causal filtering) and the minimally processed data (non-causal filtering) are presented averaged over all trials of all participants with 95% confidence intervals. In both cases, characteristic movement-related cortical potential (MRCP) patterns [20] are visible, beginning prior to movement onset. In the fully processed dataset, the waveform exhibits a phase shift. Figure 4 presents the averaged activity of channel C1 of the fully processed dataset for the rest condition, separated into active and passive rest. The averaged passive rest ERP shows lower amplitudes and overall reduced low frequency oscillations. In Figure 5, for channel C1, the EEG signals were band-pass filtered in the alpha band (8 Hz to 12 Hz) and the beta band (13 Hz to 30 Hz), squared to estimate instantaneous power and smoothed over time. The resulting power time series were z-scored within each subject and then averaged across subjects separately for movement and rest trials. Around movement onset, the movement trials exhibit a decrease in band power, followed by an increase approximately 2.5 s after movement onset. This pattern reflects the classical event-related desynchronization (ERD) followed by event-related synchronization (ERS). In contrast, the rest trials do not show such patterns. Topographic maps of EEG amplitudes be-

10.3217/978-3-99161-093-9-037

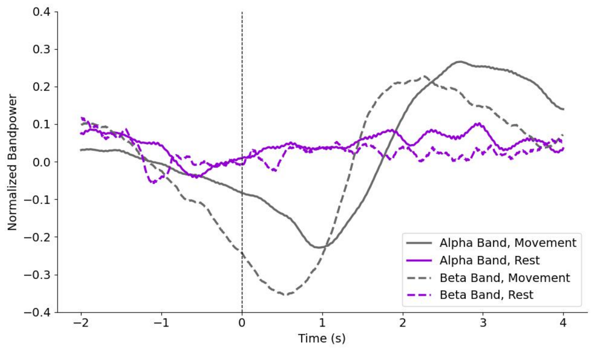

Figure 5: Time-varying bandpower of channel C1 for the alpha band (8 Hz to 12 Hz) and beta band (13 Hz to 30 Hz) averaged over movement trials (gray) and rest trials (purple). Time point 0 s marks the movement onset or virtual rest onset.

fore and after movement onset for the fully processed dataset, depicted in Figure 6, demonstrate increased central activation around movement onset and 0.5 s post-onset for movement conditions, whereas no comparable pattern is observed for rest. For visualization of the topographic maps, a Common Average Re-referencing (CAR) filter was applied, and only the subset of 31 channels present in all trials was used.

DISCUSSION

Dataset Structure: Aggregating nine experiments with 188 participants results in a large and diverse EEG dataset for hand/arm movement decoding. The unified HDF5 organization, structured by experiments, participants, and movement groups along with metadata such as detailed movement descriptions, enables flexible extraction of subsets tailored to specific research questions. Despite heterogeneous EEG channel montages, the 31-channel common subset available across all trials covers central motor areas. Moreover, domain-adaptation techniques [21] may allow the exploitation of the full channel information despite montage differences. The selected epoch length of 8 s, provides sufficient temporal context for subsequent processing steps while minimizing the risk of edge artifacts within the time window of interest.

Data Quality and Analysis: The averaged ERPs shown in Figure 3 demonstrate that the preprocessing pipeline effectively attenuated noise components, while preserving relevant signal characteristics. A clear difference in amplitudes around movement onset between the movement groups and the rest group indicate that MRCP patterns are captured in the dataset. The observed phase shift compared to the minimally processed data and the classical MRCP description [20] stems from the usage of purely causal filtering. This design choice was made to ensure compatibility with online BCI applications, where non-causal filtering is not feasible. Another aspect confirming the data quality is the difference between active and passive rest trials in Figure 4. As expected, the active rest condition exhibits increased activity, which indicates the occurrence of visual evoked potentials (VEPs)

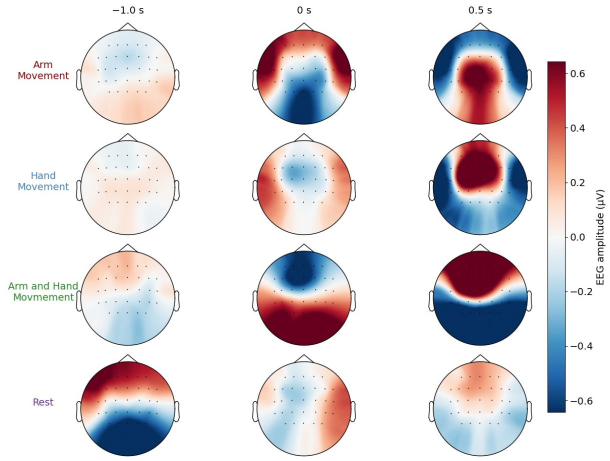

Figure 6: Topographic plots of the grand-average sensor-level EEG amplitude spatially interpolated across a 2D scalp representation of the fully processed dataset 1 s pre movement onset, during onset and 0.5 s post movement onset shown by movement group. For the plots, a CAR filter was applied. Only the subset of channels present in all trials were used.

caused by the visual cue or of contingent negative variation (CNV) signals caused by the expectancy of a motor-related task [22]. In contrast, the passive rest condition, where no cue was used, does not display such patterns. Similar components are visible in the movement conditions of Figure 3, suggesting that the observed signals reflect a combination of movement-related and cue-related activity. The bandpower differences, shown in Figure 5, further confirm the plausibility of the data. Alpha and beta band frequencies play a crucial role in movement-related tasks. ERD around movement onset, followed by ERS, are typically observed in the sensorimotor alpha and beta bands during voluntary movement [23] [24]. The topographic maps in Figure 6 further support the physiological plausibility of the data, revealing increased activation over central regions during and shortly after movement onset in movement conditions but not in rest trials.

Limitations: Despite its size and diversity, the presented dataset has certain limitations. The recordings stem from different experimental paradigms and hardware configurations, which introduces inter-experiment variability that may both be useful but also challenging for downstream tasks. Additional challenges for further machine-learning based usage of the data are the different channel montages and reference electrodes, which may be overcome by re-referencing techniques. Furthermore the dataset consists exclusively of movement execution tasks performed by able-bodied participants. While this supports transfer learning research, the generalization to movement attempts in clinical populations requires further investigation.

CONCLUSION

This work presents a curated, large-scale EEG dataset comprising nine experiments and 188 participants performing hand and arm movement execution tasks. The

10.3217/978-3-99161-093-9-037

data were uniformly processed, structured in a standardized HDF5 format, and extended with detailed metadata to support flexible data selection and transfer learning applications. Signal analyses confirm the presence of physiologically plausible movement-related patterns, while the preprocessing pipeline ensures compatibility with online BCI scenarios. The dataset is intended to facilitate the development and evaluation of generalizable deep learning models for transfer learning in EEG-based movement decoding and to support future research toward clinically applicable BCI systems. The dataset presented in this work will be publicly released in an online repository in the near future.

ACKNOWLEDGMENTS

This work was supported by the AUVA Grant "NeuroGrasp".

REFERENCES

[1] Ofner P, Schwarz A, Pereira J, Müller-Putz GR. Upper limb movements can be decoded from the time-domain of low-frequency EEG. PLOS ONE. 2017;12(8):e0182578.
[2] Ofner P, Schwarz A, Pereira J, Wyss D, Wildburger R, Müller-Putz GR. Attempted Arm and Hand Movements can be Decoded from Low-Frequency EEG from Persons with Spinal Cord Injury. Scientific Reports. 2019;9(1):7134.
[3] Motor imagery and EEG-based control of spelling devices and neuroprostheses. In: Progress in Brain Research. Elsevier, Jan. 2006, 393–409.
[4] Müller-Putz GR et al. Applying intuitive EEG-controlled grasp neuroprostheses in individuals with spinal cord injury: Preliminary results from the More-Grasp clinical feasibility study. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference. 2019;2019:5949–5955.
[5] Craik A, He Y, Contreras-Vidal JL. Deep learning for electroencephalogram (EEG) classification tasks: A review. Journal of Neural Engineering. 2019;16(3):031001.
[6] Lopes-Dias C, Sburlea AI, Müller-Putz GR. A Generic Error-related Potential Classifier Offers a Comparable Performance to a Personalized Classifier. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Jul. 2020, 2995–2998.
[7] Schwarz A, Escolano C, Montesano L, Müller-Putz GR. Analyzing and Decoding Natural Reach-and-Grasp Actions Using Gel, Water and Dry EEG Systems. Frontiers in Neuroscience. 2020;14.
[8] Crell MR, Müller-Putz GR. Handwritten character classification from EEG through continuous kinematic decoding. Computers in Biology and Medicine. 2024;182:109132.
[9] Srisrisawang N, Müller-Putz GR. Simultaneous encoding of speed, distance, and direction in discrete reaching: An EEG study. Journal of Neural Engineering. 2024;21(6):066042.
[10] Schwarz A, Ofner P, Pereira J, Sburlea AI, Müller-Putz GR. Decoding natural reach-and-grasp actions from human EEG. Journal of Neural Engineering. 2017;15(1):016005.
[11] Sburlea AI, Müller-Putz GR. Exploring representations of human grasping in neural, muscle and kinematic signals. Scientific Reports. 2018;8(1):16669.
[12] Crell MR, Kostoglou K, Sterk K, Müller-Putz GR. A novel paradigm for fast training data generation in asynchronous movement-based BCIs. Frontiers in Human Neuroscience. 2025;19.
[13] Suwandjieff P, Müller-Putz GR. EEG Analyses of visual cue effects on executed movements. Journal of Neuroscience Methods. 2024;410:110241.
[14] Egger J, Kostoglou K, Müller-Putz GR. Chrono-EEG dynamics influencing hand gesture decoding: A 10-hour study. Scientific Reports. 2024;14(1):20247.
[15] Gramfort A et al. MEG and EEG data analysis with MNE-Python. Frontiers in Neuroscience. 2013;7.
[16] Virtanen P et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods. 2020;17(3):261–272.
[17] Ablin P, Cardoso JF, Gramfort A. Faster Independent Component Analysis by Preconditioning With Hessian Approximations. IEEE Transactions on Signal Processing. 2018;66(15):4040–4049.
[18] Li A, Feitelberg J, Saini AP, Höchenberger R, Scheltienne M. MNE-ICALabel: Automatically annotating ICA components with ICLabel in Python. Journal of Open Source Software. 2022;7(76):4484.
[19] Oostenveld R, Praamstra P. The five percent electrode system for high-resolution EEG and ERP measurements. Clinical Neurophysiology. 2001;112(4):713–719.
[20] Shibasaki H, Hallett M. What is the Bereitschaftspotential? Clinical Neurophysiology. 2006;117(11):2341–2356.
[21] Zhang K et al. Application of Transfer Learning in EEG Decoding Based on Brain-Computer Interfaces: A Review. Sensors. 2020;20(21):6321.
[22] Walter WG, Cooper R, Aldridge VJ, McCALLUM WC, Winter AL. Contingent Negative Variation : An Electric Sign of Sensori-Motor Association and Expectancy in the Human Brain. Nature. 1964;203(4943):380–384.
[23] Pfurtscheller G, Silva FH Lopes da. Event-related EEG/MEG synchronization and desynchronization: Basic principles. Clinical Neurophysiology. 1999;110(11):1842–1857.
[24] Müller GR, Neuper C, Rupp R, Keinrath C, Gerner HJ, Pfurtscheller G. Event-related beta EEG changes during wrist movements induced by functional electrical stimulation of forearm muscles in man. Neuroscience Letters. 2003;340(2):143–147.