Artificial neural network based automatic detection of motor evoked potentials.

# Abstract

**Introduction:** Motor evoked potentials (MEP) can be detected using various methods that determine signal changepoints. The current detection methods perform well given a high signal to noise ratio. However performance can diminish with artefact such as those arising due poor signal quality and external electrical potentials. Part of the problem is likely because the methods ignore the morphology of a signal making it impossible to differential noise from MEPs. The lack of the knowledge of the signal morphology also limits the applications, such as identification of abnormality, of the detection methods. **Methods:** Here, we investigated a new detection method able to learn MEP morphology using artificial neural networks. To build an end-to-end MEP model, we trained a neural network model with architecture based on combined CNN and LSTM, using sample MEP data recorded from able-bodied individuals. The MEP detection capability of the classifier was compared with that of a changepoint based detection method. **Results**: The result shows that ….. **Conclusion**: …..

# Introduction

* What is MEP
* Characteristics of MEP  
  {Morphology, polarity, latency, amplitude and time width}

Figure X. MEP morphologies.

* What is MEP used for
* Why would you autodetect MEP (A SUMMARY OF THE Application SECTION IN DISCUSSIONS)
* What is the current method of autodetecting MEP
* How do we want to innovate MEP detection here.  
  Following the recent success of deep learning an obvious question is what tasks, either currently performed manually or automatically, could be improved with deep artificial neural network. Here, we investigate the utility of deep learning for automated detection of MEPs.

# Methods

## Data collection

### Participants

Seven right handed (Edinburgh handedness inventory, \cite{oldfield1971} ) mean, Y ± X, 3 females) able-bodied people, and five people with a spinal cord injury participated in this study. The demographics of the later is shown in Table 1. The participant gave their informed consents. The study was approved by the University and NHS ethics committee.

**[Table 1]** Demographics of people with a spinal cord injury.

### Spinal cord stimulation

Electrical stimulation was delivered using the Digitimer DS7. The active electrode (positive) was placed on the skin overlying a spinal spinus process with 3.2 cm round self adhesive electrode with the return electrodes (negative) placed bilaterally(split electrode) over the iliac crest using 5cm x 9cm rectangular self-adhesive electrode. The pulses are rectangular, 1 ms duration, and arrive in pairs separated by 50 ms, and was delivered at most 0.5 Hz. The stimulation amplitude depended on the resting motor threshold (RMT) of the target muscle.

### Data recording

EMG data was bilaterally acquired over the muscle bellies of biceps brachii (BB) , extensor digitorium communis (EDC) and flexor digitorium superficialis (FDS). The data was collected using the EPRecorder package and the SX230 – 1000 Biometrics LTD sensor at a sample rate of 1000 Hz with a filter factory preset between 20-460 Hz.

### Procedure

Participants were seated upright with the hands on a table, forearm pronated, and the elbow at 90o. Stimulation and recording electrodes were attached as indicated above. Participant were asked to maintains, as close as possible, an upright torso and head position, during data acquisition. Stimulation was delivered over the spinal. With the participant sitting upright with the head in the upright position, the spinal levels T1, C6, C5, C4 and C3 were identified. The location of the spinous processes, C3-C4, C4-C5, C5-C6, C6-C7, and C7-T1 were noted. Stimulation electrode was placed over C3-C4 to identify the RMT of the target muscle chose as the biceps brachii of the dominant hand. RMT was defined as the lowest stimulus intensity that elicited sMEPs of ≥50 μV peak-to-peak amplitude at the target muscle in at least 5 out of 10 consecutive trials . Starting from C3-C4, through to C7-T1, paired-pulse electrical stimulation was delivered at intensity range of 20%-160% of the RMT of the target muscle at 10% interval(3 stimuli at each intensity). During the stimulation, EMG data were recorded and saved.

## Data analysis

### Data pre-processing and labelling.

Data pre-processing was conducted using the EPRecorder which allowed visualisation, cleaning, epoching, labelling etc. The data labelling functionality was specifically developed for this project. Each recoded dataset was epoched and each epoch was visual inspected and rejected if noisy. Data epochs were also rejected if they have excessive background activity determined by comparing the statistical properties of the prestimulus against post stimulus data. Specifically, using EPRecorder, an epoch was rejected if its mean post stimulus activities is not significantly higher than that of the prestimulus activities (100 ms pre stimulus). The cleaned data was labelled to indicate a maximum of three components comprising: {background: [1 0 0], stimulus artefact: [0 1 0], response: [0 0 1] }. Figure #FIG-LABELLING shows examples of the epochs encountered and their labels visually. The labelled data was exported by concatenating all epochs and channels from each recorded dataset to obtain a column vector with accompanying three further columns indicating the labels for each sample.

[ ]

Fig #FIG-LABELLING: Examples epochs and their example labels.

Neural network architecture  
Given the different temporal characteristics and morphologies of MEP, a generalised detector must be capable recognising the wave regardless of its, 1) latency, 2) polarity, 3) morphology, 4) amplitude, and 5) time width. This is a task suited to a convolutional neural networks (CNNs) due to their translation, scale and rotation invariance properties. This implies that, e.g. if an MEP wave and its time-shifted version are each fed to a CNN, both will equally be recognised. Another type of layer that may be applicable is Long Short-Term Memory (LSTM). LSTMs are well-suited to capture dependent sequential information such as morphology, and their temporal dynamics over several time steps.  
  
Various architectures were investigated including a U-net-like setup that downsamples and then upsamples input feature along the time axis with the opposite effect along the feature axis. To investigate an architecture, the data available for training was split into a training set and a validation set. The ratio of the split was 80%:20% respectively. These dataset were used to train each architecture with sufficient epochs to cause overfitting. Using this method it was confirmed that 1D CNN layers connected to an LSTM layer lead to sufficient network learning of the dataset features. Various configurations of these two layers were further investigated. The chosen architecture based primarily on the network capacity and capability to overfit the training dataset is shown Figure Y.

[ ]  
Figure Y. Neural network architecture.

Final training and testing  
The architecture shown in Figure Y was trained for the final time by re-combining the training set and the validation set. The network was training until it overfitted the dataset with a callback on model checkpoint used to capture the best model. The best model was used to perform inference on the testing dataset.

### Deployment and comparison with standard detection method

The final classifier model was deployed for an end-user using EPRecorder. This is to showcase the practicality of the developed model. Custom made scripts were used to export the model to MATLAB environment. Due technical issues on importing 1D input models built with latest Keras version into MATLAB, the forward pass for 1D CNN, LSTM, Dense, and other necessary layers were custom written to allow frictionless deployment within the environment. The deployment allowed automated delineation of MEPs on that loaded in EPRecorder. To do this, the output of the classifier is post-processed as follows ….

For comparison, a changepoint MEP detection method was also implemented in EPRecorder. The changepoint algorithm works as follows. The area per sample and peak-to-peak of the pre-stimulus (baseline) of all included epochs are computed as features. For each of the features, a 95% confidence interval are constructed and scaled. The first and the last changepoints of the epochs are determined within post-stimulus signal. The first changepoint is taken as the possible start of a response while the last changepoint is taken as the possible stop of the response. The epoch signal between the start and stop is the possible response region. For each epoch, the are per sample and peak-to-peak of the possible response region. A response is detected for an epoch if the following criteria are satisfied: a) the area per sample and peak-to-peak of the possible response region of the epoch is greater than the upper-limit of a scaled 95% confidence interval which was computed using the pre-stimulus data, b) the peak-to-peak amplitude is greater than 0.05 mV.

Spinally evoked MEP data acquired from people with spinal cord injury (SCI) loaded were used to compare the performance of the developed model and the changepoint MEP detection method.

# Results and discussions

## Data characteristics

Recorded responses have a mean of X±Y mV peak-to-peak amplitude, x±y mV ms area, and x±y seconds width, according to the labels. The total training and validation data had XX samples amounting YY minutes of data. There was XX, YY and ZZ minutes of data for the *background*, *stimulus artefact*, and *response* labels respectively.

## Training and testing

The result of cross validation is shown in **Table #ACCU-CROSS-VAL**. The average training and testing accuracies are 90.0±0.7 and 82.6±3.4 respectively. These results may not reflect the true power of the classifier in practice given that they are based on identification of individual sample in the datasets. In practice the interest is the specific smoothed accuracy for the responses and not e.g. the sample-sample accuracy for stimulus artifacts. Therefore in practice the classifier output was post-processed to consider only the response classification with the transient scores smoothed across samples which may lead to more or less power for the classifier.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Participant | Conv-LSTM % accuracy | | Attention % accuracy | |
|  | Training | Testing | Training | Testing |
| S1 | 90 | 68 |  |  |
| S2 | 93 | 84 |  |  |
| S3 | 91 | 77 |  |  |
| S4 | 91 | 80 |  |  |
| S5 | 88 | 84 |  |  |
| S6 | 88 | 97 |  |  |
| S7 | 89 | 88 |  |  |
| **Total** | **90.0±0.7** | **82.6±3.4** |  |  |

**Table #ACCU-CROSS-VAL:** Accuracies of leave one out cross validation from the able-bodied datasets. \*, validation result. Total, mean±SEM

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Participant | Conv-LSTM % accuracy | | Attention % accuracy | |
|  | Training | Testing | Training | Testing |
| S1 | 93 | 85 |  |  |
| S2 | na | na |  |  |
| S3 | 94 | 81 |  |  |
| S4 | 95 | 83 |  |  |
| S5 | 93 | 83 |  |  |
| S6 | 92 | 97 |  |  |
| S7 | 93 | 94 |  |  |
| **Total** | **93.3±0.42** | **87.2±2.7** |  |  |

## Deployment on EPRecorder

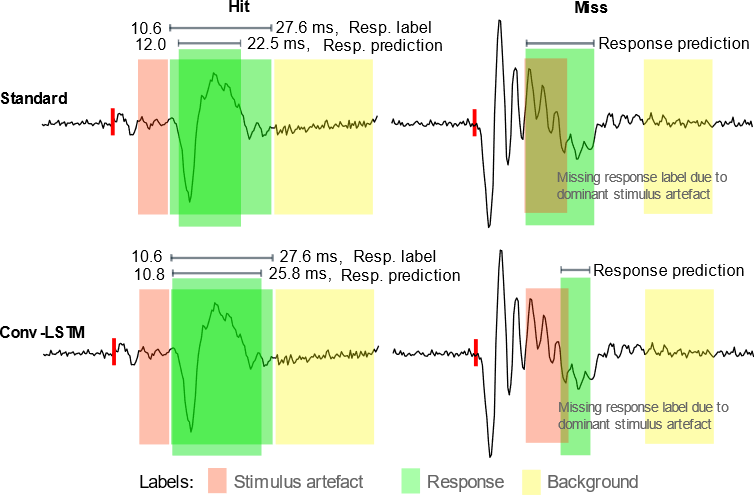
### Comparison with the standard detection method

Figure **#fig-standard-vs-conv-lstm** shows the example of a hit and a miss in EPRecorder for the standard and Conv-LSTM methods. It shows the original labels and the region of the epoch identified by the methods as containing a valid response. Both method correctly identified the presence of a response but the start and stop times of the detected response vary from that of the label. For the miss detection, typically the standard method detected a large proportion of the stimulation artefact as a valid response. On the other hand the classifier detected a small proportion of the stimulation artefact, specifically the region with a response-like depression in the waveform. These results are typical in this study; because the standard method has no knowledge of morphology of a valid waveform while the classifier make a data-informed decision tending to detect waveforms with a similar morphology as a response.

**Table #REF-DEPLOY\_RESULT** shows the summary of the classification in practice for each classification method. The classification accuracies are x±y% for the standard, x±y% for the Conv-LSTM, and x±y% for the Attention classifier.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Standard | | | Conv-LSTM | | | Attention | | |
| Parti. | #Epochs | Acc. % | Start time RMSE | Stop time RMSE | Acc. | Start time RMSE | Stop time RMSE | Acc. % | Start time RMSE | Stop time RMSE |
| S1 | 520 | 92 | 1.7 | 6.2 |  |  |  |  |  |  |
| S2 | 516 | 85 | 1.29 | 5.98 | 95 | 2.89 | 4.65 |  |  |  |
| S3 |  |  |  |  |  |  |  |  |  |  |
| S4 |  |  |  |  |  |  |  |  |  |  |
| S5 |  |  |  |  |  |  |  |  |  |  |
| S6 |  |  |  |  |  |  |  |  |  |  |
| S7 |  |  |  |  |  |  |  |  |  |  |
| T1 |  |  |  |  |  |  |  |  |  |  |
| T2 |  |  |  |  |  |  |  |  |  |  |
| T3 |  |  |  |  |  |  |  |  |  |  |
| T4 |  |  |  |  |  |  |  |  |  |  |
| T5 |  |  |  |  |  |  |  |  |  |  |

**Table #REF-DELOPLOY-RESULT**: Detection response accuracy, and start and stop time RMSE for full response detection in EPRecorder. The results for each able-bodied participant was obtained using the corresponding cross validation classifier. For the patients, the same classifier used for participant S2 was used. The start and stop rms are relative to the labels of the responses. Acc, accuracy; RMSE, root mean squared error.



**Figure #fig-standard-vs-conv-lstm**: Examples of a hit and miss for the standard and Conv-LSTM methods. The data taken from participant S2 is the same for both method. For the ‘Miss’, no label is available because the response-like waveform is judged to be excessively contaminated by the stimulation artifact.

To show how the classifier compare with the standard method as regards to start/stop time detection, the start/stop times of the for the classifier was compared with that of the standard method. The data was pooled across the able-bodied (normative) participants, with concatenation across all subjects and epochs. Using t-test the result shows that …RESULTS Of T—TEST………….

* Eamples of correctly and misclassified epochs (before and after post processing).

# Discussions

## Applications

Automated perceptual detection is useful in research and clinical practice for consistent, reliable identification and delineation of evoked, induced, and involuntary responses from e.g. brain or muscles. A specific application is in clinical electrophysiological assessment of muscles, and integrity of corticospinal pathways and their excitability as well as assessment of involuntary muscle responses. An example is in analysis of somatosensory evoked potential in intraoperative monitoring, and automated realisation of recruitment curve and averages of potentials in response to electromagnetic stimulation \cite{groppa2012practical, fournier2019clinical, gugino1990somatosensory, luck2012event}.

Another useful application is identification of deformation in evoked responses which may be due to a disorder \cite{fournier2019clinical}. This may require methods such as autoencoders to detect morphological outliers in evoked responses.

Automated detection of induced response is also useful in certain BCI application where it can be used to identify readiness potentials for e.g. paired pulse stimulation \cite{mrachacz-Kersting2012precise}.

In spinal cord stimulation, automated detection of evoked responses is useful for fast and efficient functional mapping of the spinal cord. This can be used as a neurophysiological outcome and/or for automated localisation of stimulation sites for targeted therapy by determining responses due to dorsal root stimulation, manifested in post-activation depression \cite{oh2022cervical}.

# Conclusions

# Research

* With Convolutional neural network the following issues can be resolved.
  + The wave should be detected regardless of it's temporal placement/offset.
  + Polar opposite waves should both be recognised as waves.
  + The wave should be detected regardless of it's time width. (But does convolution solve this one?)

References

\cite{groppa2012practical} Groppa, S., Oliviero, A., Eisen, A., Quartarone, A., Cohen, L.G., Mall, V., Kaelin-Lang, A., Mima, T., Rossi, S., Thickbroom, G.W. and Rossini, P.M., 2012. A practical guide to diagnostic transcranial magnetic stimulation: report of an IFCN committee. Clinical Neurophysiology, 123(5), pp.858-882.

\cite{fournier2019clinical} Fournier, E. and Tabti, N., 2019. Clinical electrophysiology of muscle diseases and episodic muscle disorders. Handbook of Clinical Neurology, 161, pp.269-280.

\cite{gugino1990somatosensory} V. Gugino and R. J. Chabot, “Somatosensory evoked potentials.,” International anesthesiology clinics, vol. 28, no. 3, pp. 154–164, 1990.

\cite{luck2012event} Luck, S. J. (2012). Event-related potentials. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), APA handbook of research methods in psychology, Vol. 1. Foundations, planning, measures, and psychometrics (pp. 523–546). American Psychological Association. https://doi.org/10.1037/13619-028

\cite{mrachacz-Kersting2012precise} N. Mrachacz-Kersting, S. R. Kristensen, I. K. Niazi, and D. Farina, “Precise temporal association between cortical potentials evoked by motor imagination and afference induces cortical plasticity.,” The Journal of Physiology, vol. 590, no. Pt 7, pp. 1669– 1682, 2012.

\cite{oh2022cervical} Oh J, Steele AG, Varghese B, Martin CA, Scheffler MS, Markley RL, Lo YK, Sayenko DG. Cervical transcutaneous spinal stimulation for spinal motor mapping. iScience. 2022 Oct 21;25(10).