Recurrent Neural Network Based Early Prediction of Future Hand Movements

Philipp Koch, Huy Phan, Marco Maass, Fabrice Katzberg, and Alfred Mertins

***Abstract*— This work focuses on a system for hand prostheses that can overcome the delay problem introduced by classical approaches while being reliable. The proposed approach based on a recurrent neural network enables us to incorporate the sequential nature of the surface electromyogram data and the proposed system can be used either for classification or early prediction of hand movements. Especially the latter is a key to a latency free steering of a prosthesis. The experiments conducted on the first three Ninapro databases reveal that the prediction up to 200 ms ahead in the future is possible without a significant drop in accuracy. Furthermore, for classification, our proposed approach outperforms the state of the art classifiers even though we used significantly shorter windows for feature extraction.**

I. INTRODUCTION

Major aims of upper limb prostheses are the recovery of most of the functionalities of a human hand as well as the simplification of the daily routine for amputees. Therefore, the hand movement recognition based on surface electromyogram (sEMG) data is a key factor for those prostheses [1], [2], [3]. To meet the requirements, an intuitive prostheses usage and a variety of different hand motions have to be enabled.

Most of the previous works on hand movement recognition focused on hand movement classification during data acquisition with a certain delay [4], [5], [6], [7]. In the standard classification scheme features are calculated for a window of sEMG data around a current time point. Given the extracted features representing the signal window, a pre-trained classifier is used to determine the corresponding hand movement. Arguably, this scheme comes with a significant drawback, the time delay. Two factors contribute to the time lag. One of them is the time necessary for data acquisition as well as feature extraction and classification. The other factor is the context window of sEMG signal around the current time point that needs to be acquired in order to perform the feature extraction. This delay is caused by the classification concept and is half of the window length in size (see Fig. 1). As Smith et al. [8] showed for this window-wise classification approaches, longer windows lead to better accuracies. Consequently, windows of a sufficient length have to be used in order to guarantee adequate results of the classification systems. But this causes a noticeable delay with possibly severe consequences for the usability of a prosthesis. So, with the classical classification methods, like Random Forest and Support Vector Machine, one also faces an unavoidable tradeoff problem between the delay and the classification accuracy.

While even more enhanced machine learning techniques like convolutional neural networks were applied [5], [9], [10], recurrent neural networks (RNNs) have been mostly ignored for determining hand movements based on sEMG signals. However RNNs such as long short-term memory (LSTM) [11] are known to be ideal for analysing sequential data. Thus, in this paper, we will investigate them for the analysis of the stream of sEMG signals in an upper limb prosthesis.

Philipp Koch, Fabrice Katzberg, and Alfred Mertins are with the Institute for Signal Processing, University of Lübeck, 23562 Lübeck, Germany {koch,katzberg,mertins}@isip.uni-luebeck.de

Huy Phan is with the Institute of Biomedical Engineering, University of Oxford, Oxford OX37DQ, United Kingdom huy.phan@eng.ox.ac.uk

Marco Maass is with the Institute for Signal Processing and the Graduate School for Computing in Medicine and Life Sciences, University of Lübeck, 23562 Lübeck, Germany maass@isip.uni-luebeck.de

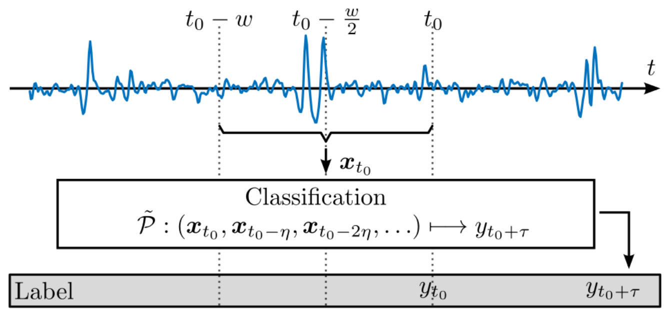

Fig. 1. Illustration of the proposed approach. Talking into account the representations of previous windows $\boldsymbol{x}_{t_0-\eta}, \boldsymbol{x}_{t_0-2\eta}$ and so on, the classifier assigns a label $y_{t_0+\tau}$ to the current window represented by $\boldsymbol{x}_{t_0}$. $\tau = -\frac{w}{2}$ is the widely used formulation of the classification problem which causes a delay of $\frac{w}{2}$. We define $\tau = 0$ as classification and for $\tau > 0$ we name it early prediction.

The work in [12] showed that it is beneficial to make use of the temporal nature of the sEMG signals for both the classification and the early prediction task. Here, early prediction describes the determination of the hand movement for some time point in the future by using information of the sEMG signal up to the current time point. The system in [12] employed random forest as the base. This basic classifier, however, is incapable of modelling temporal progression and therefore, suboptimal in capturing the sequential dynamics of sEMG signal. Therefore, in this work, an RNN based approach for both classification and early prediction of hand movements is proposed. The output of the RNN does not exclusively depend on the current input data but also on previous data of the sequence seen by the RNN. This enables an analysis of sEMG data with respect to their sequential nature. Consequently, RNNs can find patterns along the time dimension and keep track of previous information that might be helpful for improving performance for classification of hand movements without causing additional time delays by e.g. longer windows and additional feature extraction. Because the RNN provides information about the temporal context and the previous behaviour of the sEMG signal, the proposed approach is especially promising for the early prediction of hand movements for which it is necessary to deduce from the past to the future.

Since the RNN allows us to cover different degrees of temporal dependencies, we try to achieve a higher temporal resolution by reducing the length of a window from 200 ms to 100 ms without loosing performance. By using these shorter windows, the introduced time delay as well as the necessary computational resources are reduced. Another advantage of the RNNs is that the storage of previous data is handled internally. Consequently, there is no need to store data of previous time steps, which is advantageous due to the lack of computational resources in prostheses.

To validate the possibility of early prediction, we introduce an error measure that gives better information about how well the future label is determined if the current window's label is different from the future label. Experiments on the first three databases of the NinaPro project [13], including a database containing data of amputees, show that the proposed framework achieves better classification results using even significantly shorter windows than all baseline systems.

II. Reformulation of the Hand Movement Classification Problem

In the following we explain the widely used classification task and formally introduce the early prediction task. Also, the early prediction task will be adapted in a way that an early prediction with respect to the sequential nature of the sEMG data becomes possible. This means, the determination of the desired hand movement could be performed by e.g. a RNN.

A. Typical Classification System

Assume that a control system of a hand prosthesis acquires the sEMG signal window-wise. The sEMG signal window is then represented by a D-dimensional feature vector denoted as $\boldsymbol{x} \in \mathbb{R}^D$. Furthermore, let $y \in \mathbb{L}$ with $\mathbb{L} = \{1, 2, ..., C\}$ denote a label specifying one of the C possible hand movements. As soon as a new feature vector is available, a typical control system determines the corresponding hand movement using classifier $\mathcal{C}$. Formally, this pre-trained classifier performs the mapping

$$\mathcal{C} : \boldsymbol{x}_t \in \mathbb{R}^D \longmapsto y_t \in \mathbb{L}, \tag{1}$$

where t denotes the time. The label y_t corresponds to the hand movement at time t deducted from $\boldsymbol{x}_t$. This is different from many other works, where the label for $\boldsymbol{x}_t$ is $y_{t-\frac{w}{2}}$ with w being the window length. Consequently, the typically used classification approach introduces a delay of half the window length. With our formulation, this delay is avoided.

B. From Classification to Early Prediction

The objective of the classification is to determine the current hand movement given the current signal. In contrast, in early prediction the aim is to predict a future hand movement given the signal up to the current time point. Formally, the mapping becomes

$$\mathcal{P} : \boldsymbol{x}_t \in \mathbb{R}^D \longmapsto y_{t+\tau} \in \mathbb{L}, \tag{2}$$

with τ denoting the offset between the current time t and the target time. The same classifiers as in Section II-A can be used as predictors $\mathcal{P}$, however, they should be trained differently.

For $\tau > 0$ the predictors can reduce or even eliminate time delays in prosthesis steering. This offers opportunities to enhance the responsiveness and the naturalness of the prosthetic hand.

C. System for Sequence Analysis

Since sEMG signals are sequential by nature, previous signal parts should also carry information useful for dealing with the current window. The findings in [12] confirm this and show that the incorporation of earlier signal parts is beneficial for the classification accuracy as well as for the early prediction accuracy. In order to integrate information of previous windows into the classification without causing delays by lengthening the windows or additional feature extractions, a classifier should be able to determine the desired hand movement of the current window with respect to the previous ones. Therefore, we reformulate the classification and early prediction in this section. Since the classifier $\mathcal{C}$ depends on the current input as well as on information of previous inputs, the mapping becomes

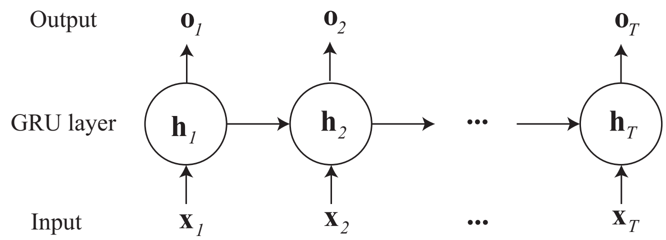

Fig. 2. Illustration of the proposed RNN architecture based on a GRU cell.

$$\tilde{\mathcal{C}} : (\boldsymbol{x}_t \in \mathbb{R}^D, \boldsymbol{x}_{t-\eta}, \boldsymbol{x}_{t-2\eta}, \cdots) \longmapsto y_t \in \mathbb{L}. \tag{3}$$

The early prediction mapping in (2) becomes

$$\tilde{\mathcal{P}} : (\boldsymbol{x}_t \in \mathbb{R}^D, \boldsymbol{x}_{t-\eta}, \boldsymbol{x}_{t-2\eta}, \cdots) \longmapsto y_{t+\tau} \in \mathbb{L}. \tag{4}$$

This formulation has significant advantages. First, it enables a classifier or early predictor to determine a hand movement of the current window represented by $\boldsymbol{x}_t$ with respect of the previous windows $\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \cdots$. Information on previous windows could be dynamically stored within the classifier itself. A class of classifiers that are capable of handling this are the RNNs. They can handle sequential input and store information within the state. With every new sample, information is added as well as deleted from the state. Second, there is no limitation from how far back information are stored.

III. Classifier and Predictor

In the following, we present the algorithm used for both the classification and the early prediction task. Although several different classification frameworks could be used for the proposed problem, the aim of this work is not to find the optimal one, but to find one that is suitable for applications with limited computational resources. Nevertheless, the used algorithm was developed with respect to possible applications, for instance upper limb prosthesis, and their lack of computational resources. We used an RNN architecture based on a gated recurrent unit (GRU) [14] in order to keep the computational effort low while still being able to make use of the sequential information.

In general, an RNN iterates over all feature vectors $\boldsymbol{x}$ of a sequence $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_T)$ and calculates for each feature vector a hidden state vector $\boldsymbol{h}$. For a time step $t \in [1, 2, ..., T]$, the corresponding hidden state vector $\boldsymbol{h}_t$ is computed using the hidden layer function $\mathcal{H}$:

$$\boldsymbol{h}_t = \mathcal{H}\left(\boldsymbol{x}_t, \boldsymbol{h}_{t-1}\right). \tag{5}$$

The hidden layer function of the chosen GRU cell can be described with the combination of the following four equations:

$$\boldsymbol{r}_t = \sigma_g\left(\boldsymbol{W}_{xr}\boldsymbol{x}_t + \boldsymbol{W}_{hr}\boldsymbol{h}_{t-1} + \boldsymbol{b}_r\right), \tag{6}$$

$$\boldsymbol{z}_t = \sigma_g\left(\boldsymbol{W}_{xz}\boldsymbol{x}_t + \boldsymbol{W}_{hz}\boldsymbol{h}_{t-1} + \boldsymbol{b}_z\right), \tag{7}$$

$$\tilde{\boldsymbol{h}}_t = \sigma_h\left(\boldsymbol{W}_{xh}\boldsymbol{x}_t + \boldsymbol{W}_{hh}\left(\boldsymbol{r}_t \odot \boldsymbol{h}_{t-1}\right) + \boldsymbol{b}_h\right), \tag{8}$$

$$\boldsymbol{h}_t = \boldsymbol{z}_t \odot \boldsymbol{h}_{t-1} + (1 - \boldsymbol{z}_t) \odot \tilde{\boldsymbol{h}}_t. \tag{9}$$

In above equations, the variables $\boldsymbol{r}$ and $\boldsymbol{z}$ denote the reset gate vector and the update gate vector, respectively. Variable $\tilde{\boldsymbol{h}}$ represents the candidate vector for the new hidden state. The weight matrices and the biases necessary to calculate $\boldsymbol{r}$, $\boldsymbol{z}$ and $\tilde{\boldsymbol{h}}$ are denoted with $\boldsymbol{W}$ and $\boldsymbol{b}$, respectively. The activation function σ_g of the gates is a sigmoid function. σ_h denotes a hyperbolic tangent function. The element-wise vector product is denoted by $\odot$.

Because for applications like upper limb prostheses the goal is to assign a class to each window, the output $\boldsymbol{o}_t$ is calculated for each time step t by

$$\boldsymbol{o}_t = \boldsymbol{W}_{hy}\boldsymbol{h}_t + \boldsymbol{b}_y. \tag{10}$$

As the algorithm should be small in order to be applicable in prostheses, we do not stack multiple GRU cells.

The output of the single cell is used as input for one fully-connected (FC) layer. This last layer of the network is used to assign a class label to the given input by approximating the probability of occurrence for each class. For a given input vector $\boldsymbol{x}$ the output of a FC layer can be described by

$$\boldsymbol{y} = \sigma\left(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}\right). \tag{11}$$

The matrix $\boldsymbol{W}$ represents the learned weights used for calculating the weighted linear combination of the input vector and $\boldsymbol{b}$ denotes a bias vector. The nonlinear activation function is symbolised by σ. Since we want to estimate a probability distribution in the final layer, the softmax function is used as activation function.

In order to be able to increase the number of training examples, we used a sequence-to-label approach. This means that the performance of the network is judged by looking at the result for the last window only.

IV. New Error Measure for True Prediction

In order to evaluate the performance of a predictor in the critical case – the true prediction case – we introduce the true prediction accuracy $\mathrm{acc_{tp}}$, which is similar to the regular accuracy acc. Let N_{tp} be the number of samples corresponding to the true prediction case and let $\hat{N}_{tp}$ denote the number of true prediction case samples, a correct label is assigned. Then, the true prediction accuracy $\mathrm{acc_{tp}}$ is computed by

$$\mathrm{acc_{tp}} = \frac{\hat{N}_{tp}}{N_{tp}}. \tag{12}$$

Especially in combination with the regular accuracy, this error measure enables the investigation of the tradeoff between the overall accuracy and the performance in the true prediction case.

V. Experiments

We evaluated the performance of a recurrent classifier for hand movement classification and compare it with standard classifiers. Furthermore, we investigated the feasibility of a recurrent early prediction system. Therefore, different prediction time offsets and their influence on the prediction performance were analysed.

A. Datasets

The experiments for the evaluation of the proposed approaches were conducted on three databases from the Ninapro project [13]. The databases are publicly accessible and have been used previously for research proposes in the field of sEMG based hand movement determination [5], [15], [16]. The proposed approach was tested for 78 subjects, 67 of them were intact and the other 11 subjects in question were trans-radial amputated.

The first database (DB I) contains sEMG signals of 53 different hand movements (including rest) for all 27 subjects who were abled and healthy. Ten repetitions of each type of movement (excluding rest) were performed by every subject. The muscular activity was captured using ten *MyoBock 13E200-50* electrodes (Otto Bock HealthCare GmbH, www.ottobock.com). The second database (DB II) includes measurements of 40 abled and healthy subjects whereas the data from 11 trans-radial amputated subjects form the third database (DB III). However, both the second and third database are slightly different from the first one regarding the experimental setup. The databases DB II and DB III contain data for 50 different hand movements (including rest). However, the subjects only performed six repetitions for each hand movement (excluding rest) instead of ten like in DB I. For data acquisition 12 *Trigno Wireless* electrodes (Delsys, Inc, www.delys.com) were used.

For a proper comparison with previous works, we followed the data preprocessing scheme, feature extraction, as well as the data splits in [13]. Specifically, the repetitions two, five (and seven in case of database one) were used for evaluation. The remaining repetitions formed the training set. In all experiments, the performance of the proposed approach was evaluated for each subject. For each database the average accuracy over all subjects is reported.

B. Preprocessing and Features

We followed a widely used processing scheme that was proposed by Englehart and Hudgins [17]. This contains preprocessing, segmentation of the signal into windows as well as a feature extraction on window level. The preprocessing includes a channel-wise normalisation of the signals in order to achieve mean-free signals with unit standard deviation. The calculation of the necessary statistics was based on the train data. Afterwards, overlapping windows of length $100\ \mathrm{ms}$ were cropped out of the normalized signal. Consecutive windows overlapped by $90\ \%$ equivalent to a hop of $10\ \mathrm{ms}$.

In order to represent a window, different features were calculated to form the feature vector characterising the given window. The chosen features are similar to the ones in [13] and were calculated for each channel individually. The used features are the root mean square as well as time domain statistics [18] such as mean absolute value, zero crossings, slope sign changes, and waveform length. Additionally, we calculated a histogram (HIST). The used histogram had 20 bins and was created given a threshold of three times the standard deviation. Consequently, HIST adds 20 features to the feature vector. Also, the results of a marginal discrete wavelet transform with a db7 wavelet of order three were included as features.

With all feature vectors being generated, we normalized each feature for every channel individually between zero and one. The necessary statistics were calculated using training data only.

C. Experimental Results

The following results were obtained by RNNs described in Section III. The used network only consists of a single GRU cell with a state size of 256. All reported results for both accuracy and true prediction accuracy are averaged across all subjects of the corresponding database.

In order to show that the proposed approach is sufficient for handling the classification task, we compare the performance of our approach using $100\ \mathrm{ms}$ long windows with previous works using windows twice as long. For DB I and II, the proposed system achieved accuracies of $79.2\ \%$ and $76.7\ \%$ respectively. With these results, the proposed system surpasses the best systems in [13] by $3.9\ \%$ and $1.4\ \%$ absolute, respectively. For DB III and thus for the amputees, our main target group, our system performs even better. We achieved an classification accuracy of $53.8\ \%$ and outperform [13] by $7.5\ \%$ absolute. For all three databases, the proposed approach improves the accuracy significantly compared

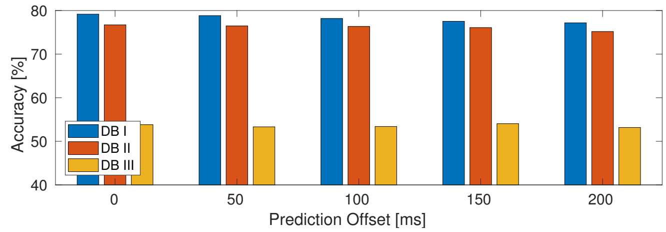

Fig. 3. Accuracy obtained with the RNN based network.

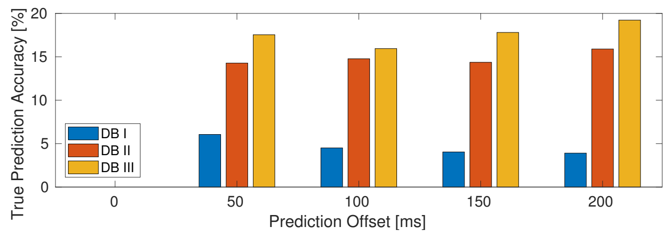

Fig. 4. True prediction accuracy obtained with the RNN based network.

with the previous works, even though the shorter windows were used. In [10] even though a window size of 200 ms was used, we outperform the reported results by 1.4 % absolute on DB I.

Since DB III contains data of amputees and is the most important case for us, we conducted the experiments with a window length of 200 ms for this database in order to show that longer windows improve the performance of our system. We achieved an accuracy of 58.6 % which is about 5 % better compared to the results for 100 ms long windows.

In the following we focus on the early prediction task. For five different prediction offsets, the corresponding results are reported in Figs. 3 and 4. The offset of 0 ms corresponds to assigning a hand movement to the end of the window. For all other offsets a future hand movement is predicted. The results in Fig. 3 show that for abled subjects like in DB I and II the accuracy decreases with increasing prediction offset. However, this is different for amputees. The results for the different early prediction offsets are similar. No clear decay can be observed. Taking the true prediction accuracies reported in Fig. 4 into account it becomes obvious that the proposed system is able to predict future hand movements for all three databases. It works especially well for the amputees. Nearly 20% of the hand movement changes are predicted correctly. The promising results can also be seen on the second database. The lack of true prediction accuracy for DB I might be due to the used electrodes and their low sampling frequency and onboard preprocessing. All in all, the results for the early prediction task are encouraging because the used databases have the distinction of containing mostly static motions with only a few changes of hand movements resulting in very little number of training examples for the true prediction case. Nevertheless, in all cases the true prediction accuracy is significantly better than guessing.

VI. Conclusion

In this work, we showed that our RNN based approach outperforms the typical window-wise approaches for hand movement classification. Especially for amputees, the proposed approach leads to a significantly improved accuracy (7.5 % and 12.2 % absolute for 100 ms and 200 ms respectively compared to those reported in [13]). This and low computational costs of a small RNN as well as the effortless handling of sequential data makes this approach very suitable for upper limb prostheses. Furthermore, the approach has shown promising performance for the early prediction task. The promising true prediction accuracy showed in this work implies the feasibility of future hand movement prediction.

References

[1] C. Castellini and P. van der Smagt, "Surface emg in advanced hand prosthetics," *Biol. Cybern.*, vol. 100, no. 1, pp. 35–47, Feb. 2009.

[2] Heather Daley, Kevin Englehart, Levi Hargrove, and Usha Kuruganti, "High density electromyography data of normally limbed and transradial amputee subjects for multifunction prosthetic control," *J. Electromyogr. Kinesiol.*, vol. 22, no. 3, pp. 478 – 484, 2012.

[3] A. H. Al-Timemy, G. Bugmann, J. Escudero, and N. Outram, "Classification of finger movements for the dexterous hand prosthesis control with surface electromyography," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 3, pp. 608–618, May 2013.

[4] Y. Zhang, G. Wang, C. Teng, Z. Sun, and J. Wang, "The analysis of hand movement distinction based on relative frequency band energy method," *BioMed Research International*, vol. 2014, Nov. 2014.

[5] M. Atzori, Cognolato M, and H. Müller, "Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands," *Front. Neurorobot.*, vol. 10, no. 9, Sep. 2016.

[6] M. P. Mobarak, R. M. Guerrero, J. M. Gutirrez Salgado, and V. L. Dorr, "Hand movement classification using transient state analysis of surface multichannel EMG signal," in *2014 Pan American Health Care Exchanges (PAHCE)*, Apr. 2014, pp. 1–6.

[7] A. Gijsberts, M. Atzori, C. Castellini, H. Müller, and B. Caputo, "The movement error rate for evaluation of machine learning methods for semg-based hand movement classification," *IEEE Trans. Neural Systems Rehab. Eng.*, vol. 22, no. 4, pp. 735 – 744, Jul. 2014.

[8] L. H. Smith, L. J. Hargrove, B. A. Lock, and T. A. Kuiken, "Determining the optimal window length for pattern recognition-based myoelectric control: Balancing the competing effects of classification error and controller delay," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 2, pp. 186–192, Mar. 2011.

[9] X. Zhai, B. Jelfs, R. H. M. Chan, and C. Tin, "Self-recalibrating surface emg pattern recognition for neuroprosthesis control based on convolutional neural network," *Front. Neurosci.*, vol. 11, no. 379, 2017.

[10] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface emg images," *Sci. Rep.*, vol. 6, no. 36571, 2016.

[11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[12] Philipp Koch, Huy Phan, Marco Maass, Fabrice Katzberg, and Alfred Mertins, "Early prediction of future hand movements using sEMG data," in *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2017, pp. 54–57.

[13] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G. Mittaz Hager, S. Elsig, G. Giatsidis, F. Bassetto, and H. Müller, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Sci. Data*, vol. 1, no. 140053, Dec. 2014.

[14] K. Cho, B. van Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," in *Proc. EMNLP*, 2014, pp. 1724–1734.

[15] Agamemnon Krasoulis, Sethu Vijayakumar, and Kianoush Nazarpour, "Evaluation of regression methods for the continuous decoding of finger movement from surface emg and accelerometry," in *Proc. International IEEE/EMBS Conference on Neural Engineering (NER)*, 2015, pp. 631–634.

[16] Meena AbdelMaseeh, Tsu-Wei Chen, and Daniel W. Stashuk, "Extraction and classification of multichannel electromyographic activation trajectories for hand movement recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 6, pp. 662–673, 2016.

[17] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control.," *IEEE Trans. Biomed. Engineering*, vol. 50, no. 7, pp. 848–854, 2003.

[18] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Engineering*, vol. 40, no. 1, pp. 82–94, Jan 1993.