2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)
4-7 March 2018
Las Vegas, Nevada, USA

Translating sEMG Signals to Continuous Hand Poses using Recurrent Neural Networks

Fernando Quivira[1], Toshiaki Koike-Akino[2], Ye Wang[2], and Deniz Erdogmus[1]

Abstract— **In this paper, we propose a hand pose estimation approach from low cost surface electromyogram (sEMG) signals using recurrent neural networks (RNN). We use the Leap Motion sensor to capture the hand joint kinematics and the Myo sensor to collect sEMG while the user is performing simple finger movements. We aim at building an accurate regression model that predicts hand joint kinematics from sEMG features. We use RNN with long short-term memory (LSTM) cells to account for the non-linear relationship between the two domains (sEMG and hand pose). Additionally, we add a Gaussian mixture model (GMM) to build a probabilistic model of hand pose given EMG data. We performed experiments across 7 users to test the performance of our approach. Our results show that for simple hand gestures such as finger flexion, the model is able to capture hand pose kinematics precisely.**

I. Introduction

Non-invasive surface electromyogram (sEMG) recordings on the forearm contain useful information for decoding muscle activity and hand kinematics [1], [2]. sEMG has been used by researchers to develop intuitive robotic prosthesis interfaces either via pattern recognition using physiological features or via classical control schemes [3], [4]. Classification approaches attempt to estimate hand posture from a pre-defined set using continuous sEMG signals. Classifiers such as support vector machines [1], linear discriminant analysis [5], artificial neural networks [6], fuzzy logic [7], Gaussian mixture models (GMM) [3], among others have been proposed using a wide variety of features (zero-crossings of raw EMG, mean absolute deviation, root mean square of the signal, etc.). However, such approaches are not valid when attempting to fully reconstruct hand kinematics.

Hand pose estimation solutions have been proposed using stereo imaging [8], tracking gloves [9], ultrasound [10], among others. Camera-based methods do not work well when there is occlusion. Moreover, some camera-based methods require geometric calibration that can be cumbersome to set up in complex environments [11]. Alternatively, glove-based methods, although reliable, can be cost prohibitive. A few studies have tackled the problem of estimating finger movement from sEMG [12]. In [13], a time-delayed neural network was used to estimate finger joint angles while [2] showed that a Gaussian process regression model outperformed the neural network approach.

In this paper, we propose a low-cost approach to build models that translate sEMG recordings to hand kinematics.

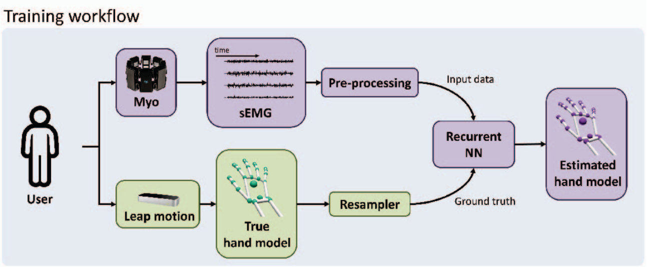


Fig. 1: Model training methodology.

The training procedure is based on inexpensive sensing with the Myo sEMG sensor and hand skeleton tracking with the Leap Motion sensor. Our approach uses recurrent neural networks (RNN) to map muscle activation signals to hand pose directly. By integrating a GMM into the RNN, we show that we can build a probabilistic model that can successfully reconstruct hand kinematics. We show our results on data from 7 users performing different gestures.

II. Methods

A. Experimental Setup

The overall data collection and analysis framework are outlined in Fig. 1. sEMG data were collected with the Myo sensor at 200 Hz. The hardware is comprised of 8 bipolar channels placed uniformly around the proximal forearm region, targeting most muscles used in hand manipulation. The Myo has been widely used as an inexpensive source of sEMG [14]. Hand pose tracking was performed with a Leap Motion sensor. The acquisition records the positions of 22 joints in the hand relative to the sensor's origin.

Users were asked to perform 7 hand gestures with 30 repetitions per gesture. Each gesture was executed in a window of 3 seconds, always returning the hand to the original resting pose. Fig. 2 shows the user interface displayed during experiments.

B. Data Acquisition

Muscle activation was estimated by computing mean absolute deviation (MAD) on a moving window over 5 samples:

$$x_{\mathrm{MAD}}(n) = \frac{1}{L} \sum_{k=n-L+1}^{n} |x(k) - m(k)|, \tag{1}$$

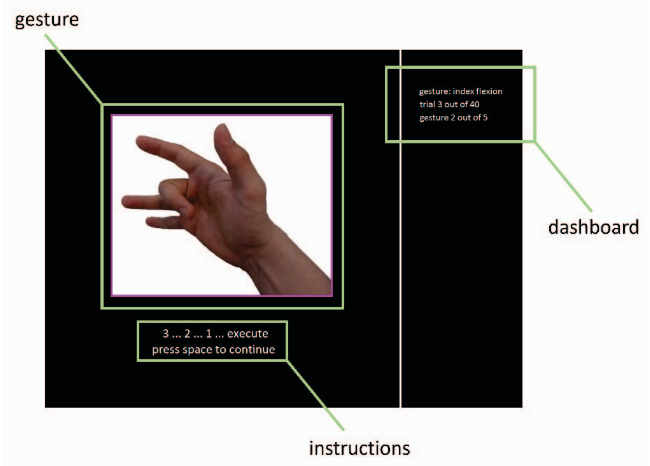
where L is the window size and $m(k)$ is the mean signal in the given window.

Hand pose data were resampled to 200 Hz using a resampling filter. Data were aligned with software triggers and

[1] Cognitive Systems Laboratory, ECE Department, Northeastern University, Boston, MA 02115, USA (e-mail: erdogmus@ece.neu.edu)

[2] Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA (e-mail: {koike, yewang}@merl.com) F. Quivira conducted this study at MERL during internship.

978-1-5386-2405-0/18/$31.00 ©2018 IEEE



gesture: index flexion
trial 3 out of 40
gesture 2 out of 5

3 ... 2 ... 1 ... execute
press space to continue

Fig. 2: Data collection instruction interface.

filter group delay was taken into account with a temporal shift. The joint coordinates were transformed to be relative with respect to the hand itself, thus making them translation and rotation invariant. Additionally, the dimensionality of the hand pose vector (22 features for each coordinate) was reduced with principal component analysis (PCA), keeping enough features to reconstruct original poses with 95% fidelity (no loss in visual quality). Finally, the temporal regression task was set to map 8 sEMG inputs (MAD processed) to 10 hand pose features. The low dimensional representation can be used to reconstruct the joint positions in the original coordinate system.

C. Regression with Recurrent Neural Networks

RNNs are expressive models that can learn not only long-term dependencies in sequential data such as time series, but also can model complex non-linear dynamics [15]. RNNs learn hidden representations $\mathbf{h}_n$ of temporal data:

$$\mathbf{h}_n = \tanh(\mathbf{W}\,\mathbf{h}_{n-1} + \mathbf{V}\,\mathbf{x}_n), \tag{2}$$

where $\mathbf{W}$ is the recurrent weight matrix and $\mathbf{V}$ is the projection to the input representation. The hidden state $\mathbf{h}$ is then used to make a prediction of the output (with a linear mapping, for example). Moreover, RNN modules like this can be stacked to learn more complex representations, yielding richer models [15]:

$$\mathbf{h}^l_n = \sigma(\mathbf{W}_l\,\mathbf{h}^l_{n-1} + \mathbf{V}_l\,\mathbf{h}^{l-1}_n), \tag{3}$$

where σ is the logistic sigmoid function and l denotes the layer index. The last layer hidden state will be used to compute an estimate of the desired output. However, training such RNNs can be difficult due to the vanishing and exploding gradient problem.

Long short-term memory (LSTM) networks are RNN modules that address the vanishing gradient problem by using gating functions that control the state dynamics [15]. At each step, an LSTM cell keeps an external output vector $\mathbf{h}$ and a hidden internal state memory vector $\mathbf{C}$ which is responsible for updating itself and the output. The computations are defined as follows:

$$\begin{aligned}
\mathbf{g}^i &= \sigma(\mathbf{W}^i\,\mathbf{h}_{n-1} + \mathbf{V}^i\,\mathbf{x}_n),\\
\mathbf{g}^f &= \sigma(\mathbf{W}^f\,\mathbf{h}_{n-1} + \mathbf{V}^f\,\mathbf{x}_n),\\
\mathbf{g}^o &= \sigma(\mathbf{W}^o\,\mathbf{h}_{n-1} + \mathbf{V}^o\,\mathbf{x}_n),\\
\hat{\mathbf{C}}_n &= \tanh(\mathbf{W}^c\,\mathbf{h}_{n-1} + \mathbf{V}^c\,\mathbf{x}_n),\\
\mathbf{C}_k &= \mathbf{g}^f \odot \mathbf{C}_{k-1} + \mathbf{g}^i \odot \hat{\mathbf{C}}_n,\\
\mathbf{h}_n &= \mathbf{g}^o \odot \tanh(\mathbf{C}_{n-1}),
\end{aligned} \tag{4}$$

where $\mathbf{W}^*$ and $\mathbf{V}^*$ are state and input projection weights. Each LSTM cell can be stacked as described earlier. In our case, the input vector was the MAD-processed sEMG signals and the output was the PCA-processed hand pose.

D. Generative Modeling with Mixture Density Networks

From a probabilistic perspective, using mean squared error as the training loss corresponds to assuming that the output can be modeled with unit variance Gaussian distribution [16]. Such model fails to capture complex relationships between output features which are likely to be present in data as complex as hand movement. Recurrent mixture density networks (RMDN) [15] have been shown to be effective in a wide variety of tasks such as hand writing modeling, trajectory prediction, among others. With this approach the output of the network is modeled probabilistically as follows:

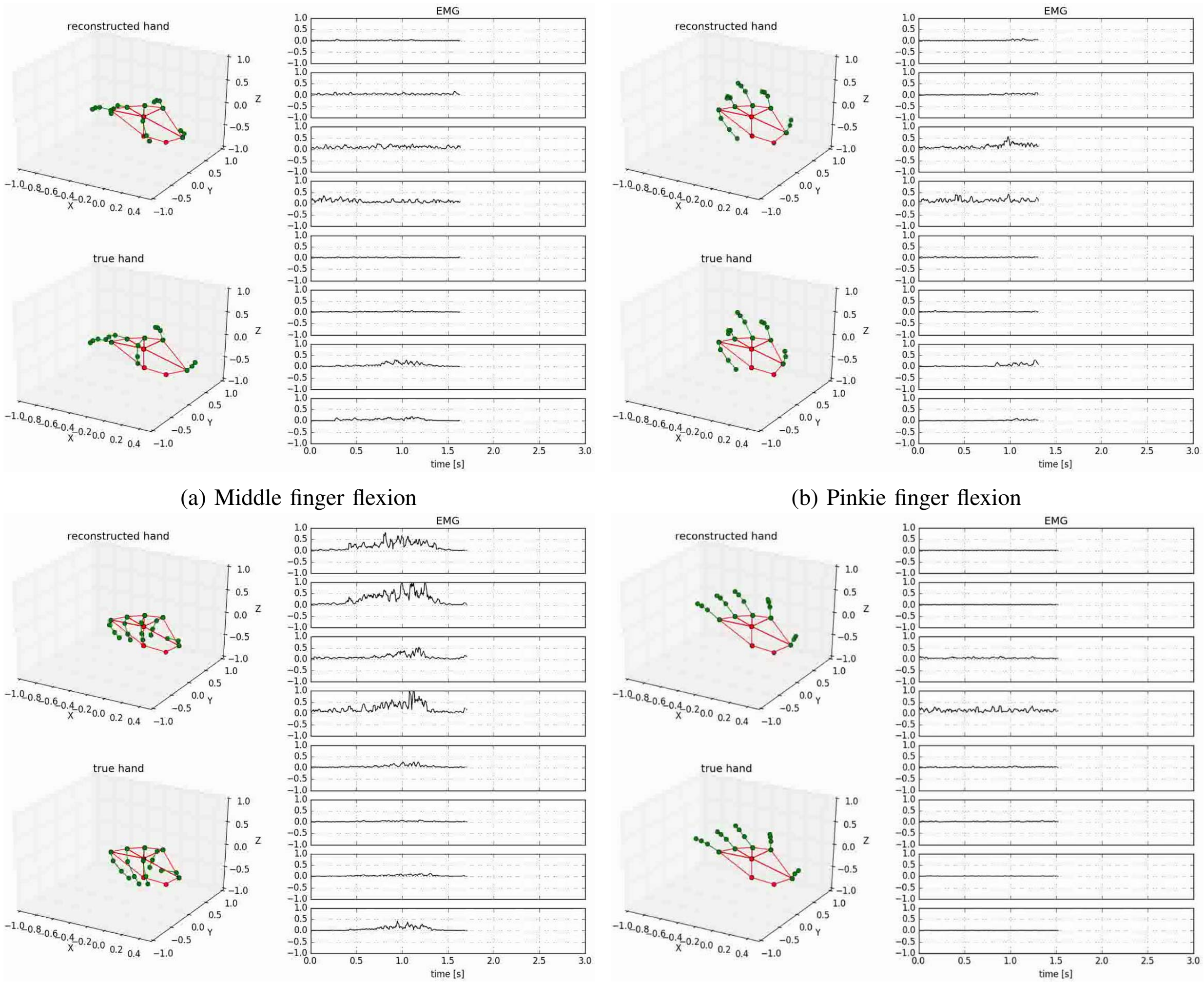
$$\begin{aligned}
p(\mathbf{y}_n|\mathbf{x}_{\leq n}) &= p(\mathbf{y}_n|\mathbf{x}_{<n}, \mathbf{x}_n)\\
&\approx p(\mathbf{y}_n|\mathbf{h}_{n-1}, \mathbf{x}_n)\\
&= \frac{1}{K}\sum_{k=1}^{K} \mathcal{N}(\mathbf{y}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\pi_k,
\end{aligned} \tag{5}$$

where $\boldsymbol{\mu}_k := \boldsymbol{\mu}_k(\mathbf{h}_{n-1}, \mathbf{x}_n)$, $\boldsymbol{\Sigma}_k := \boldsymbol{\Sigma}_k(\mathbf{h}_{n-1}, \mathbf{x}_n)$, and $\pi_k := \pi_k(\mathbf{h}_{n-1}, \mathbf{x}_n)$ are the outputs of neural networks with the previous state and the current sEMG instance x_n as inputs, and K corresponds to the number of mixtures of Gaussian distribution $\mathcal{N}(\cdot)$. This model makes the assumption that previous sEMG activity $\mathbf{x}_{<n}$ can be summarized into hidden state $\mathbf{h}_{n-1}$. To learn the parameters of this model, we minimize the negative log-likelihood function defined by:

$$L(\Theta) = \sum_{n=1}^{N} \log \sum_{k=1}^{K} \mathcal{N}(\mathbf{y}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\pi_k, \tag{6}$$

with Θ as our set of parameters including GMM parameters. This log-likelihood is optimized using truncated backpropagation through time. Since it is a composition of continuous functions (e.g. linear transformations and element-wise non-linearity) and LSTM modules, we can compute the gradients easily with automatic differentiation tools. Similar to [15], we defined our GMM parameters as follows:

$$\begin{aligned}
\mu_k &= W^{\mu}_k h^o(\mathbf{h}_{n-1}, \mathbf{x}_n) + b^{\mu}_k,\\
\boldsymbol{\Sigma}_k &= \mathrm{diag}(\sigma^2_{k,1}, ..., \sigma_{k,L}),\\
\sigma^2_{k,l} &= \mathrm{elu}(W^{\sigma}_k h^o(\mathbf{h}_{n-1}, \mathbf{x}_n) + b^{\sigma}_k) + 1,\\
\pi_k &= \frac{\exp(W^{\pi}_k h^o(\mathbf{h}_{n-1}, \mathbf{x}_n) + b^{\pi}_k)}{\sum_{j=1}^{K} \exp(W^{\pi}_j h^o(\mathbf{h}_{n-1}, \mathbf{x}_n) + b^{\pi}_j)},
\end{aligned} \tag{7}$$



(a) Middle finger flexion

(b) Pinkie finger flexion

(c) Power spherical grasp

(d) Resting pose

Fig. 3: Snapshot results from RNN-based regression, showing reconstructed and true poses as well as sEMG signal.

where $h^o(\mathbf{h}_{n-1}, \mathbf{x}_n)$ is the output of a multi-layered LSTM network for the current time step n. The covariance matrix was assumed to be diagonal to reduce the number of parameters to be learned. The GMM weights are outputs of a softmax activation to ensure proper normalization [15].

One advantage of using probabilistic models for regression tasks over deterministic ones is the capability of estimating a confidence range over the output given a series of inputs [16]. Additionally, the RMDN approach allows us to learn multimodal representations of possible hand poses [16]. In other words, with similar sEMG inputs, the model can predict more than one plausible hand pose instead of simply averaging them; multiple model tracking can realistically account for similar muscle activations yielding different hand postures.

III. Results and Discussion

We use the network architecture as follows: 5 layers with 50 LSTM cells per layer. Data aggregation was used by jittering the input/output pairs by 50 samples (0.25 seconds). This prevented the network from simply memorizing the sequence. Truncated backpropagation through time of length 5 was used for training. Gradient clipping and ℓ_2 weight decay was used to regularize network weights. 10% of the data were used for validation. 500 epochs were used to train with early stopping on a validation set. Mean squared error was used as the loss function. 5-fold cross validation was used for performance assessment. The network was trained and validated for each individual user using 80% of the data; the rest were used to estimate performance.

7 subjects were asked to perform the experiment with basic hand postures: flexion for each finger, resting state, and spherical power grasp. 40 trials per hand posture were collected for each user. Each trial lasted for 3 seconds plus 3 seconds for resting. Users were asked to perform the gesture starting from resting position. Fig. 3 shows example reconstructions from the RNN for given time instances.

Fig. 4 shows the test root mean-square error in mm for the 2 models: deterministic (Det-RNN) and probabilistic RMDN. In all users, the probabilistic model outperforms the deterministic one. RMDN used 10 mixtures of Gaussian. The estimated output for RMDN was obtained by finding the mean of the best mixture as follows: $\hat{\mathbf{y}}_n = \boldsymbol{\mu}_{k^\star}$ with

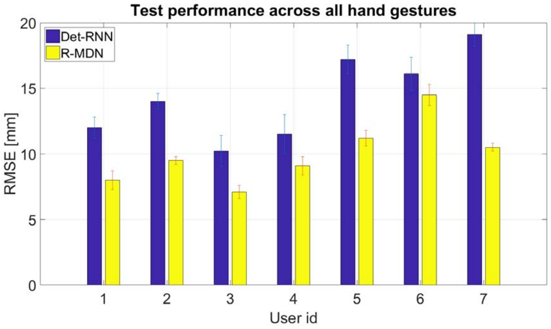


Fig. 4: RMSE of joint distances across all gestures.

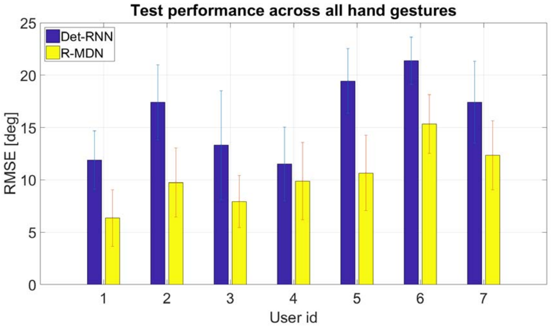


Fig. 5: RMSE of joint angles across all gestures.

$k^\star = \arg\max_k \mathcal{N}(\mathbf{y}_k|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\pi_k$. Fig. 5 shows the per-user average performance in estimating angle between finger joints. On average, RMDN outperforms the deterministic case. Fig. 6 shows the performance of thumb flexion regression, which was the lowest performing gesture in our dataset. Since the abductor policis longus is close to the wrist, the Myo forearm sensor is not able to capture isolated movements of the thumb.

The current subjects have been limited to healthy, able-bodied subjects to test the feasibility of our approach. Moreover, only simple gestures were recorded due to sensing limitations. More complex kinematics could be captured with gloves or hand markers.

IV. Conclusions

Our study shows successful reconstruction of finger movement from low-cost sEMG recordings. In our simple training scheme, RNNs are shown to be powerful time-series models that can capture the hand kinematic variability. The use of regression methods offers the potential advantage of generalizing to novel movements due to the continuous nature of the model. Further verification of our results with more complex hand postures is required to further evaluate our approach. Future work includes but is not limited to adding more complex gestures during training, using other probabilistic models such as recurrent variational autoencoders, and using more accurate sensing methods.

References

[1] M. Yoshikawa, M. Mikawa, and K. Tanaka, "Hand pose estimation using EMG signals," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2007, pp. 4830–4833.

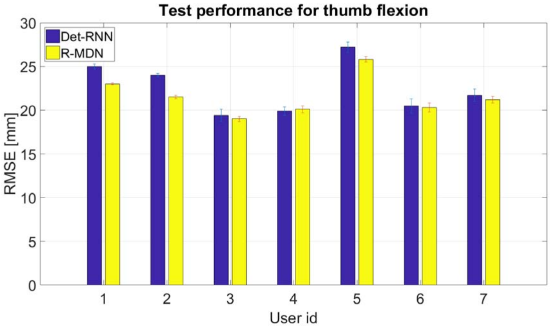


Fig. 6: RMSE of joint distances for thumb flexion.

[2] Jimson G. Ngeo, Tomoya Tamei, and Tomohiro Shibata, "Continuous and simultaneous estimation of finger kinematics using inputs from an EMG-to-muscle activation model," *Journal of NeuroEngineering and Rehabilitation*, vol. 11, no. 1, pp. 122, Aug 2014.

[3] Y. Huang, K. Englehart, B. Hudgins, and A. D. C. Chan, "A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses," *IEEE Trans Biomed Eng*, vol. 52, 2005.

[4] Ali H Al-Timemy, Rami N Khushaba, Guido Bugmann, and Javier Escudero, "Improving the performance against force variation of EMG controlled multifunctional upper-limb prostheses for transradial amputees," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 6, pp. 650–661, 2016.

[5] S. Negi, Y. Kumar, and V. M. Mishra, "Feature extraction and classification for EMG signals using linear discriminant analysis," in *2016 2nd International Conference on Advances in Computing, Communication, Automation (ICACCA) (Fall)*, Sept 2016, pp. 1–6.

[6] F. E. R. Mattioli, E. A. Lamounier, A. Cardoso, A. B. Soares, and A. O. Andrade, "Classification of EMG signals using artificial neural networks for virtual hand prosthesis control," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2011, pp. 7254–7257.

[7] M. S. Çelik and İ. Eminoğlu, "Fuzzy logic based classification for multifunctional upper limb prostheses," in *2016 Medical Technologies National Congress (TIPTEKNO)*, Oct 2016, pp. 1–4.

[8] Hee-Sung Kim, G. Kurillo, and R. Bajcsy, "Hand tracking and motion detection from the sequence of stereo color image frames," in *2008 IEEE International Conference on Industrial Technology*, April 2008, pp. 1–6.

[9] J. H. Kim, N. D. Thang, and T. S. Kim, "3-D hand motion tracking and gesture recognition using a data glove," in *2009 IEEE International Symposium on Industrial Electronics*, July 2009, pp. 1013–1018.

[10] N. Hettiarachchi, Z. Ju, and H. Liu, "A new wearable ultrasound muscle activity sensing system for dexterous prosthetic control," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, Oct 2015, pp. 1415–1420.

[11] Y. Fang, N. Hettiarachchi, D. Zhou, and H. Liu, "Multi-modal sensing techniques for interfacing hand prostheses: A review," *IEEE Sensors Journal*, vol. 15, no. 11, pp. 6065–6076, Nov 2015.

[12] M. Xiloyannis, C. Gavriel, A. A. C. Thomik, and A. A. Faisal, "Gaussian process autoregression for simultaneous proportional multi-modal prosthetic control with natural hand kinematics," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1785–1801, Oct 2017.

[13] M. Hioki and H. Kawasaki, "Estimation of finger joint angles from sEMG using a neural network including time delay factor and recurrent structure," *Int Scholarly Res Netw (ISRN) Rehabil*, vol. 2012, 2012.

[14] S. Rawat, S. Vats, and P. Kumar, "Evaluating and exploring the Myo armband," in *2016 International Conference System Modeling Advancement in Research Trends (SMART)*, Nov 2016, pp. 115–120.

[15] Alex Graves, "Generating sequences with recurrent neural networks," *CoRR*, vol. abs/1308.0850, 2013.

[16] Christopher M. Bishop, "Mixture density networks," Tech. Rep., 1994.