2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Can Deep Synthesis of EMG Overcome the Geometric Growth of Training Data Required to Recognize Multiarticulate Motions?*

Alexander E. Olsson[1], Nebojša Malešević[1], Anders Björkman[2], and Christian Antfolk[1]

***Abstract*— By being predicated on supervised machine learning, pattern recognition approaches to myoelectric prosthesis control require electromyography (EMG) training data collected concurrently with every detectable motion. Within this framework, calibration protocols for simultaneous control of multifunctional prosthetic hands rapidly become prohibitively long—the number of unique motions grows geometrically with the number of controllable degrees of freedom (DoFs). This paper proposes a technique intended to circumvent this combinatorial explosion. Using EMG windows from 1-DoF motions as input and EMG windows from 2-DoF motions as targets, we train generative deep learning models to *synthesize* EMG windows appertaining to multi-DoF motions. Once trained, such models can be used to complete datasets consisting of only 1-DoF motions, enabling simple calibration protocols with durations that scale linearly with the number of DoFs. We evaluated synthetic EMG produced in this way via a classification task using a database of forearm surface EMG collected during 1-DoF and 2-DoF motions. Multi-output classifiers were trained on either (I) real data from 1-DoF and 2-DoF motions, (II) real data from only 1-DoF motions, or (III) real data from 1-DoF motions appended with synthetic EMG from 2-DoF motions. When tested on data containing all possible motions, classifiers trained on synthetic-appended data (III) significantly outperformed classifiers trained on 1-DoF real data (II), although significantly underperformed classifiers trained on both 1- and 2-DoF real data (I) ($p < 0.05$). These findings suggest that it is feasible to model EMG concurrent with multiarticulate motions as nonlinear combinations of EMG from constituent 1-DoF motions, and that such modelling can be harnessed to synthesize realistic training data.**

I. INTRODUCTION

In order to restore upper limb functionality, a myoelectric prosthetic hand should ideally require little effort to control and at the same time allow the user to perform a wide range of grasps and motions. The first desideratum—naturalness of control—is in theory straightforwardly achievable by pursuing myoelectric control based on *pattern recognition* [1]. Within this framework, the prosthesis control problem is formulated as one of *statistical prediction*: by utilizing data comprised of electromyography (EMG) signals and concurrent motion labels, machine learning algorithms can be trained to *classify* EMG time segments as belonging to one out of multiple predefined motion classes. Decisions from algorithms trained in this way can be interpreted as motion commands to be sent to a motorized prosthesis.

By construction, pattern recognition myoelectric control requires EMG training data from every motion that is to be performable by the prosthesis user. This presents an obvious obstacle to achieving diversity of possible motions. The problem is exacerbated by the fact that *simultaneous control*—here meaning the ability to independently steer every kinematic degree of freedom (DoF) at the same time—is arguably one of the ultimate goals of the prosthesis control field [2]. To realize the scope of the problem of achieving simultaneous control with pattern recognition methods, one can consider the case of a prosthetic hand with D DoFs, each of which can assume S states. The number of unique motions M such a system can perform is $M = S^D$, i.e. the number of unique motions grows geometrically with the number of controllable DoFs. Thus, as the mechanical sophistication of prostheses increases, the number of motions that need to be recorded from amputees is subject to a combinatorial explosion. Corollarily, exhaustively recording EMG signals from every possible motion entails long and complicated acquisition protocols even for relatively small values of D and S.

In light of the difficulties in procuring sufficient data for training simultaneous control methods within the pattern recognition paradigm, we propose a method intended to circumvent the need for exhaustive, user-specific datasets. In brief, we introduce a novel artificial neural network (ANN) architecture and use it to explicitly model user-independent relationships between EMG from motions incorporating multiple DoFs and EMG from constituent 1-DoF motions. Once trained on pairings of EMG windows from a multi-subject dataset comprised of 1-DoF motions and all of their possible combination motions, such models can be used to *synthesize* user-specific EMG windows associated with multiarticulate motions from real examples of 1-DoF EMG windows. Using this framework, new prosthesis users would only need to perform all relevant 1-DoF motions (i.e. a total of $D \cdot S$ unique motions) for the purpose of calibration, after which standard pattern recognition control interfaces can be trained on a dataset including synthetic multi-DoF EMG. In this study, the quality of synthetic EMG produced with this method was evaluated via a classification task: performance metrics obtained from multi-output classifiers trained on real data was compared to metrics obtained from classifiers trained on real data augmented with synthetic EMG.

*Research supported by the Promobilia Foundation, the Crafoord Foundation, the European Commission under the DeTOP project (LEIT-ICT-24-2015, GA #687905), and the Swedish Research Council (DNR 2019-05601).

[1]A. E. Olsson, N. Malešević, and C. Antfolk are with the Dept. of Biomedical Engineering, Faculty of Engineering, Lund University. P.O. Box 118, Lund, Sweden. {alexander.olsson, nebojsa.malesevic, christian.antfolk}@bme.lth.se

[2]A. Björkman is with the Department of Hand Surgery, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Sahlgrenska University Hospital, Gothenburg, Sweden. anders.bjorkman@gu.se

Previous work has been successful in synthesizing EMG signals for the purpose of training data augmentation, either via explicit simulations [3] or via deep generative learning [4]. Although studies concerning EMG motion decoding has often conceptualized motions as composed of 1-DoF 'basis' motions for the purpose of multi-output classification (e.g. [2], [5]), we are not aware of any existing attempts to leverage this combinatorial view in order to synthesize EMG.

II. METHODS

A. Data

The database of EMG recordings with synchronized motion annotations used here was originally collected for the purpose of a previous study [6]. In brief, myoelectric signals were recorded from 20 healthy subjects using a Myo armband (Thalmic labs, Canada) consisting of 8 circularly arranged bipolar surface electrodes. The armband was placed enclosing the dominant forearm of the subject at a level approximately 1/3 of the distance from the elbow to the wrist. EMG signals were sampled at a rate of 200 Hz. The collection protocol entailed the use of two DoFs: (I) flexion/extension of the wrist and (II) flexion/extension of all digits simultaneously. Motions were encoded using a ternary scheme, wherein each DoF could assume 1 out of 3 values at each time point: -1 (flex), 0 (neutral/stall), or 1 (extend). This scheme allows for $S^D = 3^2 = 9$ possible motions (listed in table I), all of which were recorded. Each motion was repeated $R = 3$ times, with repetitions lasting for 5 seconds and separated by 3 s of rest.

An inter-subject leave-one-out cross-validation design was used to partition the data for the purpose of training and evaluating the generative framework. At each iteration, data from 19 subjects were used to train a synthesizer model via the procedures outlined in the sections following. Data from the remaining, held-out subject were lastly used as the basis for intra-subject classifier training and testing for the purpose of evaluating the impact adding synthetic training data has on pattern recognition performance.

TABLE I: The 9 recorded motions comprising the database.

Motion Description	Ternary Encoding
Rest	[0, 0]
Flexion of the wrist	[-1, 0]
Extension of the wrist	[1, 0]
Flexion of the digits	[0, -1]
Extension of the digits	[0, 1]
Flexion of the wrist & Flexion of the digits	[-1, -1]
Flexion of the wrist & Extension of the digits	[-1, 1]
Extension of the wrist & Flexion of the digits	[1, -1]
Extension of the wrist & Extension of the digits	[1, 1]

B. Preparation of Synthesizer Training Data

At each cross-validation iteration, raw EMG time series $X_c[n]$ from the 19 subjects constituting the training data were clipped at the 1:st and 99:th percentiles and normalized to the range $[-1, 1]$ separately for each channel $c = 1, ..., 8$. Conditioned signals $S_c[n]$ obtained in this way were, separately for each subject, segmented into a set of EMG time windows $\{\mathbf{E}^r_{i,j}\}$, where $\mathbf{E}^r_{i,j} \in \mathbb{R}^{600\times 8}$ represents the middle 3 s (600 samples) of the r:th repetition of the motion with ternary encoding $i\in\{-1, 0, 1\}$, $j\in\{-1, 0, 1\}$. With 8 unique nonrest motions (from table I) repeated 3 times, $3 \cdot 8 = 24$ such time windows were obtained per subject. To construct training data for the ANN model, EMG window instances originating from 2-DoF motions—designated as the *target* value—were paired with two EMG window instances originating from its two constituent, 1-DoF movements—designated as *input* values. By including every possible pairing of repetitions, this approach resulted in a training set consisting of a total of $M \cdot R^3 = 4 \cdot 3^3 = 108$ input-output pairs per subject, where $M = S^D - D - 1 = 4$ is the number of unique 2-DoF motions and $R = 3$ is the total number of available repetitions for each motion.

C. Synthesizer Model

All deep learning models used in this study were implemented and instantiated using the TensorFlow 1.12 library executed with Python 3.6. Inspired by the *variational autoencoding* [7] approach to distribution modeling, the synthesizer model architecture introduced and applied here (illustrated graphically in fig. 1) performs a mapping from two EMG windows recorded during two different single-DoF motions—DoF 1 and DoF 2, respectively—to an EMG window recorded concurrently with the motion consisting of DoF 1 and DoF 2 active simultaneously. Specifically, the architecture consists of 3 modules: two encoder networks, a mixer network, and a decoder network.

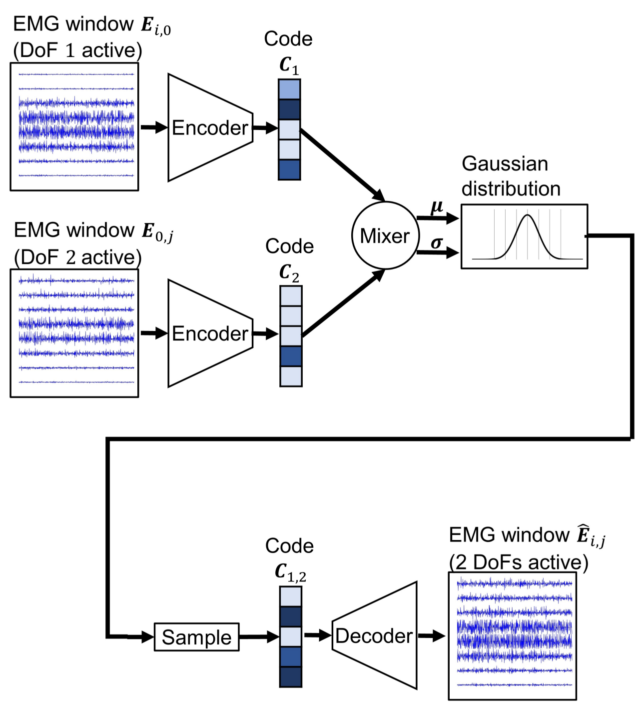


Fig. 1: Schematic overview of the synthesizer model.

- **Encoder**. The encoder networks transform the input 1-DoF EMG windows $\mathbf{E}_{i,0} \in \mathbb{R}^{600\times 8}$ and $\mathbf{E}_{0,j} \in \mathbb{R}^{600\times 8}$ into code vectors $\mathbf{C}_1 \in \mathbb{R}^{1024}$ and $\mathbf{C}_2 \in \mathbb{R}^{1024}$, respectively. They consist of 6 2D convolution layers with kernel sizes $[9\times 1]$, $[1\times 3]$, $[15\times 1]$, $[1\times 3]$, $[15\times 1]$, and $[1\times 3]$; output depths 64, 64, 256, 256, 1024, and 1024; and strides $[6\times 1]$, $[1\times 2]$, $[10\times 1]$, $[1\times 2]$, $[10\times 1]$, and $[1\times 2]$. Each layer is followed by leaky ReLU activation, layer normalization, and 50% dropout. The encoder networks share weights during training.
- **Mixer**. The mixer network transforms the encoder output codes $\mathbf{C}_1$ and $\mathbf{C}_2$ into a mean vector $\boldsymbol{\mu} \in \mathbb{R}^{1024}$ and standard deviation vector $\boldsymbol{\sigma} \in \mathbb{R}^{1024}$. This is achieved using a single fully connected layer of output size 2048 with linear activation function, whose output is split in two. The obtained vectors are used to define a 1024-dimensional distribution $N(\boldsymbol{\mu}, \text{diag}(\boldsymbol{\sigma}))$, from which a sample code $\mathbf{C}_{1,2}$ is drawn and presented as output.
- **Decoder**. The decoder network transforms an input mixture code $\mathbf{C}_{1,2} \in \mathbb{R}^{1024}$ into an 2-DoF synthetic EMG time window $\hat{\mathbf{E}}_{i,j} \in \mathbb{R}^{600\times 8}$. Mirroring the architectures of the encoders, the decoder consists of 6 2D transposed convolution layers with kernel sizes $[1\times 3]$, $[15\times 1]$, $[1\times 3]$, $[15\times 1]$, $[1\times 3]$, and $[9\times 1]$; output depths 1024, 256, 256, 64, 64, and 1; and strides $[1\times 2]$, $[10\times 1]$, $[1\times 2]$, $[10\times 1]$, $[1\times 2]$, and $[6\times 1]$. The first 5 transposed convolution layer are followed by leaky ReLU activation, layer normalization, and 50% dropout; the final layer is followed by a linear activation.

Models were trained end-to-end to minimize the loss $\mathcal{L}$:

$$\mathcal{L} = \mathcal{L}_r + \mathcal{L}_s + \mathcal{L}_d \tag{1}$$

The *reconstruction* loss $\mathcal{L}_r$ quantifies the discrepancy between the synthetic 2-DoF EMG window produced by the network and the target 2-DoF EMG window provided during training. It is obtained by computing the squared Euclidean distance between the absolute frequency spectrum of the target $\mathbf{E_{i,j}}$ and the absolute frequency spectrum of the decoder output $\hat{\mathbf{E}}_{i,j}$:

$$\mathcal{L}_r = \| |FFT(\mathbf{E}_{i,j})| - |FFT(\hat{\mathbf{E}}_{i,j})| \|_2^2 \tag{2}$$

The *spread* loss $\mathcal{L}_s$ incentivizes the encoder network to learn different code representations for EMG windows originating from different 1-DoF motions:

$$\mathcal{L}_s = \max(0, 10 - \|\mathbf{C}_1 - \mathbf{C_2}\|_2^2) \tag{3}$$

The *divergence* loss $\mathcal{L}_d$ regularizes the ANN model by penalizing mixer output distributions $N(\boldsymbol{\mu}, \text{diag}(\boldsymbol{\sigma}))$ with large Kullback–Leibler divergence compared to the normal distribution with zero mean and unit variance:

$$\mathcal{L}_d = \text{KL}(N(\boldsymbol{\mu}, \boldsymbol{\sigma}), N(\mathbf{0}, \mathbf{1})) \tag{4}$$

Loss minimization was performed iteratively by use of the Adam algorithm [8] with $\eta = 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, training in mini-batches of size 1024 for a total of 5000 epochs. At the start of the training, all network parameters were given initial values via Glorot initialization.

D. Synthetic Data Evaluation

Signals from the held-out subject were split into classifier training and testing data on the basis of repetition: data from the first and second repetition of each motion for training and data from the last repetition for testing. This designation was used to produce 4 pattern recognition training datasets:

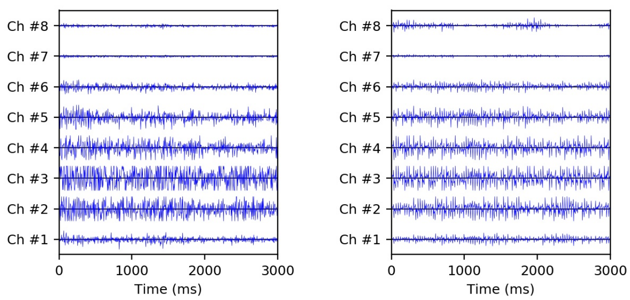
- **I: Complete**. Real EMG from all motions.
- **II: Pruned**. Real EMG from all 1-DoF motions.
- **III: Deep Augmentation**. Real EMG from 1-DoF motions appended with synthetic EMG from 2-DoF motions produced by the trained synthesizer model. This training set was created by feeding every combination of 1-DoF motion repetitions as inputs to the network; as the first two repetitions of each motion had been selected for pattern recognition training, this resulted in $2^2 = 4$ synthetic EMG repetitions per unique 2-DoF motion.
- **IV: Additive Augmentation**. Real EMG from all 1-DoF motions appended with synthetic EMG repetitions for all 2-DoF motions created by simply summing every possible pair of real EMG signals from 1-DoF motion repetitions. This reference training set was included as a comparison to ensure that any classifier performance gain brought about by the deep synthesizer model reflects properties of the synthesized signals and not simply an inflated training set.

For each of the 4 datasets, signals were segmented into feature vectors with a sliding window technique using 250 ms (50 samples) time windows with step size 5 ms (1 sample). A conventional time-domain feature set [9] (mean-absolute value (MAV), zero-crossings (ZC), slope-sign change (SSC), waveform length (WL)) was extracted from each window, producing a 32-dimensional feature vector at each window location. A target 2-dimensional ternary label vector associated with each feature vector was created by a DoF-wise majority vote over the samples of the time window. All 32 features were normalized to have zero mean and unit variance across each pattern recognition training set. A single pattern recognition test set was created in an identical manner using the last repetition of all 8 nonrest motions.

For each of the 4 pattern recognition training sets, one 3-class Linear Discriminant Analysis (LDA) classifier was trained per DoF; the $D = 2$ classifiers trained on each training set can together be viewed as a single multi-output, multi-class classifier whose performance is indicative of the quality of the training set. A choice of two multi-output classification metrics—Exact Match Rate (EMR) and Hamming Loss(HL)—were computed to quantify the performance of each of the 4 multi-output classifiers when operating on the test set. For both metrics, the difference in mean (across the 20 cross-validation iterations) was computed between the pruned (II) and deep augmentation (III) datasets, between the full (I) and deep augmentation (III) dataset, and between the deep augmentation (III) and additive augmentation (IV) datasets. Wilcoxon signed-rank tests with $\alpha = 0.05$ were used to assess the significance of observed differences.

III. RESULTS

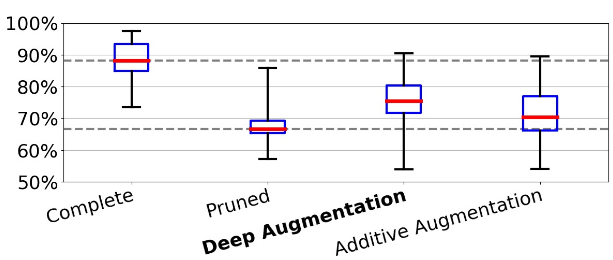
An example of a synthetic EMG window produced by a deep synthesizer model and an example of a corresponding real EMG window (i.e. from the same subject and motion) are shown together in fig. 2. Performance metrics achieved by LDA classifiers trained on complete (dataset I), pruned (dataset II), and partially synthetic (datasets III and IV) data are summarized in fig. 3. For the EMR metric, the mean increase between dataset II (67.86%) and III (75.42%) was 7.59% ($p = 0.00039$), the mean increase between dataset IV (70.60%) and dataset III was 4.82% ($p = 0.0064$), and the mean increase between dataset III and dataset I (87.96%) was 12.54% ($p = 0.000089$). Similarly for the HL metric, the mean decrease between dataset II (17.95%) and dataset III (13.41%) was 4.54% ($p = 0.00078$), the mean decrease between dataset IV (15.72%) and dataset III was 2.31% ($p = 0.014$), and the mean decrease between dataset III and dataset I (6.24%) was 7.17% ($p = 0.000088$).



(a) Real EMG window. (b) Synthetic EMG window.

Fig. 2: Example of real and synthetic EMG windows from the same 2-DoF motion and subject.

(a) Exact Match Rate.



(b) Hamming Loss.

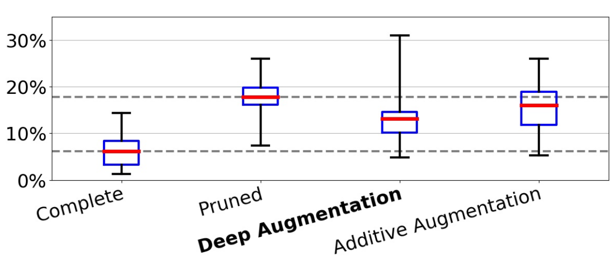


Fig. 3: Multi-output performance metrics of LDA classifiers trained on datasets I-IV.

IV. CONCLUSION

This paper has proposed a deep learning-based approach for completing EMG pattern recognition training datasets containing only 1-DoF motions with synthetic EMG associated with multi-DoF motions. To assess the viability of the approach, the impact on classifier performance of partially synthesized data was investigated. The inclusion of EMG synthesized via the novel approach resulted in significantly higher performance (as measured by two metrics) compared to when using unaugmented data and augmented data created with a naive reference method. Nevertheless, the quality of the synthesized 2-DoF data was found to be strictly inferior to real EMG signals from 2-DoF motions for the purpose of training pattern recognition algorithms. Even so, these findings show that it is possible to model a user-independent relationship between EMG from multi-DoF motions and EMG from constituent 1-DoF motions, and that such models can be used to generate practically applicable training data.

Although this work only involved models operating on EMG from 2-DoF motions, there is no fundamental obstacle to extending the approach presented here to an arbitrary number of controllable DoFs. However, the need to record all combinations for the purpose of initially training a synthesizer model puts a practical upper limit on the number of DoFs manageable by the approach. In addition to extending the method to dataset with additional DoFs, future work could evaluate the use of alternative machine learning methods aimed at image or signal combining (e.g. [10]).

REFERENCES

[1] E. Scheme and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use," *J Rehabil Res Dev*, 2011, vol 48, pp. 443-59.

[2] A. Krasoulis and K. Nazarpour, "Myoelectric digit action decoding with multi-output, multi-class classification: an offline analysis," *Sci Rep*, vol. 10, no. 1, 2020.

[3] A. Furui, H. Hayashi, G. Nakamura, T. Chin, and T. Tsuji, "An artificial EMG generation model based on signal-dependent noise and related application to motion classification," *PLoS one*, vol. 12, no. 6, 2017.

[4] E. Campbell, J. A. D. Cameron, and E. Scheme, "Feasibility of data-driven EMG signal generation using a deep generative model," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2020, pp. 3755-3758.

[5] A. E. Olsson, P. Sager, E. Andersson, A. Björkman, N. Malesevic, and C. Antfolk, "Extraction of multi-labelled movement information from the raw HD-EMG image with time-domain depth," *Sci Rep*, vol. 9, no. 1, 2019.

[6] A. E. Olsson, N. Malesevic, A. Björkman, and C. Antfolk, "Learning regularized representations of categorically labelled surface EMG enables simultaneous and proportional myoelectric control," *J Neuroeng Rehabil*, vol. 18, no. 35, 2021.

[7] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in *2nd International Conference on Learning Representations (ICLR2014)*, 2014.

[8] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", Preprint available: http://arxiv.org/abs/1412.6980, 2014.

[9] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunctional myoelectric control," *IEEE Trans Biomed Eng*, vol. 40, no. 1, pp. 82-94, 1993.

[10] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," Preprint available: https://arxiv.org/abs/1508.06576, 2016.