33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

# Gesture recognition in upper-limb prosthetics: A viability study using Dynamic Time Warping and gyroscopes

Konstantinos Dermitzakis, Alejandro Hernandez Arieta and Rolf Pfeifer

*Abstract*— **One of the significant challenges in the upper-limb-prosthetics research field is to identify appropriate interfaces that utilize the full potential of current state-of-the-art neuroprostheses. As the new generation of such prostheses paces towards approximating the human physiological performance in terms of movement dexterity and sensory feedback, it is clear that current non-invasive interfaces are still severely limited. Surface electromyography, the interface ubiquitously used in the field, is riddled with several shortcomings. Gesture recognition, an interface pervasively used in wearables and mobile devices, shows a strong potential as a non-invasive upper-limb prosthetic interface. This study aims at showcasing its potential in the field by using gyroscope sensors. To this end, we (1) explore the viability of Dynamic Time Warping as a classification method for upper-limb prosthetics and (2) look for appropriate sensor locations on the body. Results indicate an optimal classification rate of $97.53\%, \sigma = 8.74$ using a sensor located proximal to the endpoint performing a gesture.**

## I. INTRODUCTION

The requirements of (1) manipulation dexterity, (2) sensory feedback and (3) anthropomorphism are pervasive in the upper-limb-prosthetics field. The past few years have seen advances that provide satisfactory solutions to the above requirements. Several current state-of-the-art robotic hands can approximate the physiological performance and visual aspects of the human hand. However, as these advanced neuro-prostheses close the gap with their biological counterparts, a bottleneck on the communication interface between the prosthesis and its user is increasingly apparent. The task of interfacing users with dexterous prostheses is difficult as it requires either large-bandwidth communication interfaces to and from the user, or controllers capable of compensating for the lack thereof [1].

Human-prosthesis interfaces can be separated into (1) invasive and (2) non-invasive. The former gather signals directly from the user's nervous system, either via brain implants or via the surgical use of electrodes. Despite their ability to deliver high quality signals, they involve surgery and are associated with sterility and rejection issues [1]. The latter are popularly used as they do not involve surgical procedures and thus have no associated physiological setbacks. Surface electromyography (sEMG) is the standard non-invasive interface for controlling upper limb prosthetic devices. Muscle activation potentials are gathered by electrodes placed on the skin. These potentials can then be used to control a prosthesis, usually by classifying the user's intention into control commands. However, sEMG comes bundled with several problems: (1) inter-participant variability (subcutaneous fat layer thickness, forearm dimensions), (2) arm posture dependence, (3) electrode displacement, (4) muscle fatigue [2]–[4], (5) unreliable non-linear methods requiring a large amount of training data for an accurate recognition rate [5], [6], (6) electronic equipment noise and EM radiation [3], [4] and (7) a limited number of classifiable motions (bandwidth-restricted channel) [7], [8].

It is therefore apparent that research into alternative interfaces is needed. While Inertial Measurement Units (IMUs) are ubiquitously used for gesture recognition in hand-held devices [9], [10], and are just making their appearance in rehabilitation [11], there is little to no attention from the upper-limb research field. We argue that gesture recognition can serve as either (1) a viable alternative or (2) a complement to existing interfaces used to control an upper-limb prosthesis, and specifically sEMG classification.

To test the viability of controlling an upper-limb prosthesis using gesture recognition[1], we use a recently-resurfaced algorithm, dynamic time warping (DTW) [12], [13]. As DTW is very effective for personalized gesture recognition with limited training data, inter-participant variability, sensor displacement and algorithmic complexity are not an issue. Further, by utilizing gyroscopes, we are ensuring that muscle fatigue effects and equipment noise minimally affect the signal space. The only potential limitation of the algorithm is an upper limit on the number of gestures it can effectively classify that needs to be tested. Further, if DTW is to be used as an upper-limb prosthetic interface, optimal sensor placement locations should be identified. Our hypothesis is that recognition degradation will be apparent as the sensor is placed further away from the endpoint performing a gesture.

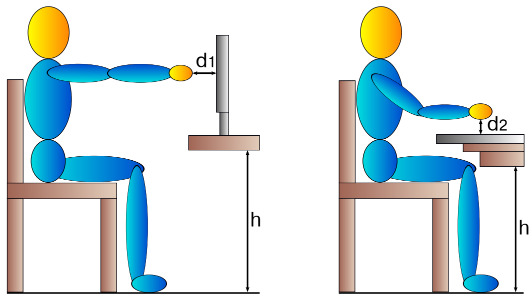
## II. EXPERIMENTAL DESIGN

### A. Dynamic Time Warping (DTW)

Dynamic time warping is a dynamic programming algorithm, matching two time-series with temporal dynamics [13]. Given time series $L = \{L_i : i \in M\}$ and $S = \{S_j : j \in N\}$ with $L_i, S_j$ three dimensional angular speed vectors, DTW provides a matching cost, calculated by traversing a so-called *warp-path* between them. For a detailed description of the algorithm see [12]. Consider a matrix of distances $D$ between time samples $L_i$, $S_j$, where $D(i, j) = \|L_i - S_j\|$. If $C(i, j)$ the cumulative cost along

K. Dermitzakis, Hernandez Arieta and R. Pfeifer are with the Artificial Intelligence Lab, University of Zurich, Switzerland dermitza@ifi.uzh.ch

[1]By gesture in this paper we mean free-space hand movements in 3D space. Such movements are, for this definition, usually preceded and followed by a non-movement period.




(a) XZ plane setup (b) XY plane setup

Fig. 1. Seating arrangement for the two plane conditions, XZ and XY. The height $h$ of the screen was adjusted for each participant individually, while the endpoint (closed fist) distances to the screen were set to $d_1 = 10$cm, $d_2 = 5$cm.

an optimal warp-path between the two time series starting from $(L_1, S_1)$ and leading to $(L_M, S_N)$, then

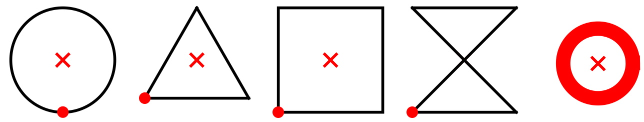
$$C(i,j) = min(C(i-1,j), C(i-1,j-1), C(i,j-1)) + D(i,j) \tag{1}$$

Additionally, optimal matching costs are normalized by the average number of points between the two time series, $C_{norm}(i,j) = 2C(i,j)/(N+M)$. To use DTW as a classifier, a library of gestures is selected, against which subsequent time-series are compared. Matching is performed by identifying the library index of the minimum matching cost of a time-series, $idx = IDX(MIN(C_k))$, where $k$ is the library index of a gesture, and $C_k$ the normalized optimal cost between library gesture $k$ and a corresponding gesture $S$ to be tested.

## B. Gestures

To identify whether DTW has a restrictive limitation on the number of gestures it can classify, we used a selection of 22 distinct gestures. They are divided into two groups, categorized as artificial ***(AG)*** and natural ***(NG)*** gestures. The first group consists of simple 2D shapes (e.g. circle, triangle and square) modified from the group in [10]. The second group consists of complex, natural motions from the Wolf Motor Test [14]. The entire set of twenty-two gestures is performed on two planes, ***XZ*** and ***XY*** (Fig. 1) and two movement conditions: free-arm ***(NS)*** and with a donned prosthetic socket ***(S)*** created by Uniklinik Balgrist, Zurich and modified for right-hand use by non-amputees. The socket with the prosthesis weighs 1141g, and imposes a motion constraint on all DOFs of the wearer's wrist, including pronation and supination occurring at the distal radioulnar joint. A tendon-based prosthetic hand [15] is attached at the distal end, configured to a closed fist.

*1) Artificial gestures:* The artificial gestures are displayed as the first four primitives shown in Fig. 2. Each primitive defines a set of two gestures: a clockwise and counter-clockwise motion traversing the gesture path[2], with the red circle defining the starting point. They collectively define a total of 8 gestures per plane.



(a) Circle (b) Triangle (c) Square (d) Hourglass (e) NG

Fig. 2. Starting condition of gesture primitives used in this study. The red disc on the natural gestures (NG) primitive is provided as timing information to the participants.

*2) Natural gestures:* The fifth primitive shown in Fig. 2 defines two gesture sets. The first set consists of an endpoint movement from mouth to target and reverse (***MT***). The second consists of an endpoint movement from rest (arm is resting vertically with the forearm held horizontally, parallel to the right femur when sitting) to target, and reverse (***RT***). They define a total of six gestures (4 for the XZ plane and 2 for the XY plane). The ***RT*** gesture set was not used on the XY plane due to the close proximity of the rest and target areas.

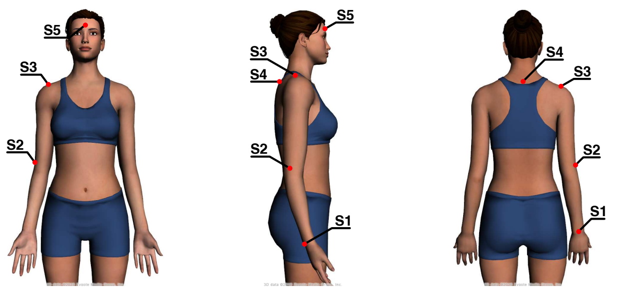
## C. Experimental procedure

Six right-handed, healthy participants (three women and three men) with an average age of 29.3 years ($\sigma = 3.39$, range $26-35$) took part in this study. All participants provided their informed consent under the approval of the Swiss Ethics Committee before the experiment.

Three body regions of interest were identified: the right upper extremity, the upper back and the head. Looking for representative body segment movements during gestures, five locations were chosen for the sensors (S1-S5): (1) the *wrist*, (2) the *lateral epicondyle area* of the humerus, (3) the *supra acromion bursa* (region of attachment between scapula and clavicle), (4) the *T1 vertebrae* and (5) the *forehead*, as seen in Fig. 3. Sensor 1 measures forearm movement and sensor 2 upper arm movement. Sensors 3-5 were used to identify whether respective scapular, upper-back and head movements are involved in upper limb motions. The sensors used are the LPR530AL (pitch and roll) and LY530ALH (yaw) gyroscopes, having a scale of $\pm 300°$/s. Sensors placed in the wrist and forehead were secured using a self-adhesive Velcro-style band. The remaining sensors were secured using double-sided adhesive tape. Data was acquired using a NI USB-6255 data acquisition device and NI LabVIEW, with a sampling frequency of 100Hz.

Following sensor placement, participants sat in front of a screen displaying the gestures to be performed, with no object in their peripersonal space constraining their motion (Fig. 1). A User Interface (UI) program was developed in Java for displaying the gestures to be performed on a screen. The participants tracked a red circle moving along the corresponding gesture path on the screen (where applicable) using their right arm. Ten trials per gesture were performed,

[2]For the hourglass primitive, we define clockwise motion as the path taken by traversing the diagonal first, and counter-clockwise motion the path taken by traversing the lower horizontal first.




(a) Front view (b) Side view (c) Rear view

Fig. 3. Gyroscope sensor placement. Five gyroscope sensors were used (S1-5). Body visualization is ©Google.

TABLE I

DTW RECOGNITION RATES (%) FOR EACH SENSOR (S1-S5) AND EACH MOTION CONDITION (NS, S)

| | NS | | | S | | |
|---|---|---|---|---|---|---|
| | $\tilde{x}$ | $\bar{x}$ | $\sigma$ | $\tilde{x}$ | $\bar{x}$ | $\sigma$ |
| S1 | 100 | 94.98 | 12.89 | 100 | 93.54 | 14.23 |
| S2 | 100 | 97.53 | 8.74 | 100 | 96.04 | 10.39 |
| S3 | 100 | 90.16 | 19.12 | 100 | 89.46 | 19.42 |
| S4 | 88.89 | 81.93 | 22.79 | 88.89 | 80.44 | 24.29 |
| S5 | 66.67 | 64.20 | 28.21 | 66.67 | 60.40 | 30.01 |

with all gestures having a cycle frequency of 1.5Hz with an inter-trial rest period of 1s. The gestures were performed without the socket, on the XZ plane followed by gestures on the XY plane (Fig. 1(b)). Once the complete set of 22 gestures was performed, there was a five minute break, after which the socket was donned on their right arm. They then performed the gesture set with the socket, first on the XY plane and then on the XZ plane. The total experiment time was one hour per participant.

## III. Data Analysis

The Matlab statistics toolbox was used for our statistical analysis. Data was first pre-processed using a two-pole butterworth, 25Hz, zero-delay, low-pass filter. Following, DTW was used to calculate recognition rates for each gesture against a library consisting of a single sample of all 22 gestures. As we are interested in individual participant recognition, gesture samples from a participant are used to provide libraries and test samples for that participant. To further improve the statistical significance of our analysis, bootstrap estimation was used [16]. For each participant, each movement condition and each sensor, for test $i, i$ = 1–10, we use the $i^{th}$ trial of each gesture to build a library consisting of 22 gestures. The remaining trials of that participant are classified against this library, resulting in a total of 600 tests. Each test produces a confusion matrix of DTW recognition rates, with the diagonal indicating the actual recognition of each gesture, providing a total of 1320 samples per sensor, per movement condition.

Our analysis consists of identifying statistically significant differences for two categories: (1) socket vs. no socket and (2) between-sensor recognition rates. As our sample distributions are not normal (Kologorov-Smirnov (KS) test, $p < 0.05$), the Quade test is used on the averaged recognition rates over all trials and gestures. Averaged distributions are similar except sensor pair 3-4 for the NS condition (KS test, $p = 0.02597$). To account for possible $\alpha$ accumulation of pairwise comparisons, we select a new minimum significance level $\alpha' = 1 - (1-\alpha)^{(1/m)}$ where $m$ is the number of tests and $\alpha$ the employed level of significance, $\alpha = 0.05$.

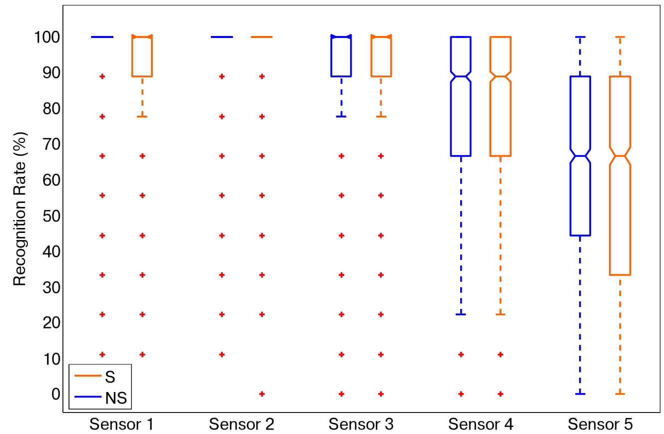



Fig. 4. DTW recognition rates by sensor and movement condition. Outliers (crosses) are points further than $\pm 2.7\sigma$ from the mean. Whiskers extend to the most extreme value not considered an outlier.

## IV. Results and Discussion

Table I shows the median ($\tilde{x}$), mean ($\bar{x}$) and standard deviation ($\sigma$) of the recognition rates for each sensor and each motion condition. A box plot of this data is shown in Fig. 4. Results for the (1) socket vs. no socket and (2) between-sensor recognition rate comparisons are presented below.

### A. No socket vs. socket

For each of the five sensors, we identify no statistically significant differences between the NS and S conditions, with the smallest $p = 0.1187$ identified on S1. The lack of significant differences between the two movement conditions is a desired phenomenon. It indicates that the altered motion dynamics and enforced motion constraints present with the donned socket do not influence the DTW recognition rates for our gesture set.

### B. Sensor differences

For the NS condition, statistically significant differences are found between sensor pairs S1-2 ($p = 0.005934$) and S2-3 ($p = 0.005934$). The differences in pair S4-5 are marginally insignificant ($p = 0.0254$) while on pair S3-4 are not significant ($p = 0.2863$). For the S condition, significant changes are found only between sensor pair S4-5 ($p = 0.005934$). The difference in pair S2-3 is marginally insignificant ($p = 0.02615$), and in pairs S1-2 and S3-4 not significant ($p = 0.2863$ for both pairs).

The differences found between sensor pairs for the NS condition reinforce our sensor placement hypothesis. In gross terms, DTW recognition rates drop when moving away from the endpoint performing a gesture. This is not however the case for sensor pair S1-2, where S2 seems to provide better recognition rates. We cannot, at this moment, attribute this observation to any effect; further experimentation is needed.

For the S condition, large variability is present in all sensors, which would explain the lack of statistically significant differences. Recognition rate variability also seems to increase as we move away from the endpoint performing a gesture. At this point, we cannot fully identify the source of this variability. A known phenomenon in neuroscientific literature, *optimal feedback control* [17], where joints distal to an endpoint movement display an increase in position variability could potentially be the source. We should also consider the simplest explanation; the head and trunk are simply not consistently contributing to our gesture set, i.e. they are not actively used.

### C. Applicability to upper-limb prosthetics

DTW classification using a gyroscope is a viable method for controlling an upper-limb prosthesis. For the library chosen, recognition rates for the upper-arm sensors are very high, both with and without the socket. While current sEMG systems can guarantee reliable classification of up to six classes [7], the number of gestures used in this study showcases the potential of the DTW algorithm in this context.

A second advantage over sEMG systems is apparent from the natural gesture sets used. Using DTW classification, controlling a prosthesis to grasp an object would stem as a direct consequence of a contextual user behaviour, e.g. an intuitive reaching motion such as the rest-to-target (*RT*) gesture. In contrast, to achieve a similar result using sEMG, a multiplicity of sensors would have to be employed, with a large volume of training data needed and questionable classification rates.

As such, we propose to either (1) implement DTW classification as a stand-alone control scheme, substituting sEMG classification, or (2) create a hybrid control system incorporating a multitude of interfaces. For some tasks, actuating the hand while keeping the limb still is required (e.g. typing). For these types of tasks, a hybrid sEMG and DTW classificator could be better suited. At the same time, such a hybrid controller could see a potential increase in classification rates, difficult to obtain with a sEMG interface alone.

## V. Conclusion

In this work we presented DTW gesture recognition in the context of upper-limb prosthetics using gyroscope sensors. As a substitute to sEMG classification, it not only effectively negates problems commonly encountered, but also serves as an intuitive human-prosthesis interface. High recognition rates and direct access to contextual information, in conjunction to minimal computational costs and minimal application-dependent issues make it a key method that is yet to be fully exploited in upper-limb prosthetic research.


## Acknowledgments

The authors would like to thank Dr. Bertolt Meyer from the Social Psychology Department, UZH, and Dr. Heike Vallery from the Sensory-Motor Systems Lab, ETHZ for their valuable input. This work is supported by the Swiss National Science Foundation project CR23I2 132702/1.


## References


[1] G. Di Pino, E. Guglielmelli, and P. M. Rossini, “Neuroplasticity in amputees: main implications on bidirectional interfacing of cybernetic hand prostheses.” *Progress in neurobiology*, vol. 88, no. 2, pp. 114–26, Jun 2009.

[2] C. Castellini and P. v. Der Smagt, “Surface EMG in advanced hand prosthetics.” *Biological cybernetics*, vol. 100, no. 1, pp. 35–47, Jan 2009.

[3] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, “Techniques of EMG signal analysis: detection, processing, classification and applications.” *Biological procedures online*, vol. 8, no. 1, pp. 11–35, Jan 2006.

[4] M. Ahsan, M. Ibrahimy, and O. Khalifa, “EMG signal classification for human computer interaction: A review,” *European Journal of Scientific Research*, vol. 33, no. 3, pp. 480–501, 2009.

[5] M. Asghari Oskoei and H. Hu, “Myoelectric control systems-a survey,” *Biomedical Signal Processing and Control*, vol. 2, no. 4, pp. 275–294, 2007.

[6] H. Yokoi, A. Arieta, R. Katoh, W. Yu, I. Watanabe, and M. Maruishi, “Mutual adaptation in a prosthetics application,” *Embodied Artificial Intelligence*, vol. 3139, pp. 629–629, 2004.

[7] C. Cipriani, F. Zaccone, S. Micera, and M. Carrozza, “On the shared control of an EMG-controlled prosthetic hand: analysis of user-prosthesis interaction,” *IEEE Transactions on Robotics*, vol. 24, no. 1, pp. 170–184, 2008.

[8] P. v. d. Smagt, M. Grebenstein, H. Urbanek, N. Fligge, M. Strohmayr, G. Stillfried, J. Parrish, and A. Gustus, “Robotics of human movements.” *Journal of physiology, Paris*, vol. 103, no. 3-5, pp. 119–32, 2009.

[9] J. O. Wobbrock, A. D. Wilson, and Y. Li, “Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes,” in *Proceedings of the 20th annual ACM symposium on User interface software and technology*, ser. UIST ’07. New York, NY, USA: ACM, 2007, pp. 159–168.

[10] J. Kela, P. Korpip, J. Mntyjrvi, S. Kallio, G. Savino, L. Jozzo, and S. Marca, “Accelerometer-based gesture control for a design environment,” *Personal Ubiquitous Comput.*, vol. 10, pp. 285–299, July 2006.

[11] X. Zhang, X. Chen, W.-H. Wang, J.-H. Yang, V. Lantz, and K.-q. Wang, “Hand gesture recognition and virtual game control based on 3d accelerometer and EMG sensors,” in *Proceedings of the 14th international conference on Intelligent user interfaces*, ser. IUI ’09. New York, NY, USA: ACM, 2009, pp. 401–406.

[12] C. S. Myers and L. R. Rabiner, “A comparative study of several dynamic time-warping algorithms for connected word recognition,” *The Bell System Technical Journal*, vol. 60, no. 7, pp. 1389–1409, 1981.

[13] H. Sakoe and S. Chiba, “Dynamic programming algorithm optimization for spoken word recognition,” *Readings in speech recognition*, pp. 159–165, 1990.

[14] S. L. Wolf, D. E. Lecraw, L. A. Barton, and B. B. Jann, “Forced use of hemiplegic upper extremities to reverse the effect of learned nonuse among chronic stroke and head-injured patients,” *Experimental Neurology*, vol. 104, no. 2, pp. 125 – 132, 1989.

[15] K. Dermitzakis and A. Hernandez Arieta, “Anthropomimetic approach to the design of a prosthetic robot hand,” in *Workshop on Understanding the Human Hand for Advancing Robotic Manipulation (RSS 2009)*, 2009, extended abstract.

[16] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley-Interscience, Nov. 2001.

[17] E. Todorov and M. I. Jordan, “Optimal feedback control as a theory of motor coordination,” *Nature Neuroscience*, vol. 5, pp. 1226–1235, 2002.