Intent Prediction Based on Biomechanical Coordination of EMG and Vision-Filtered Gaze for End-Point Control of an Arm Prosthesis

Nili E. Krausz, Denys Lamotte, Iason Batzianoulis, Levi J. Hargrove, Silvestro Micera, and Aude Billard

***Abstract*— We propose a novel controller for powered prosthetic arms, where fused EMG and gaze data predict the desired end-point for a full arm prosthesis, which could drive the forward motion of individual joints. We recorded EMG, gaze, and motion-tracking during pick-and-place trials with 7 able-bodied subjects. Subjects positioned an object above a random target on a virtual interface, each completing around 600 trials. On average across all trials and subjects gaze preceded EMG and followed a repeatable pattern that allowed for prediction. A computer vision algorithm was used to extract the initial and target fixations and estimate the target position in 2D space. Two SVRs were trained with EMG data to predict the x- and y- position of the hand; results showed that the y-estimate was significantly better than the x-estimate. The EMG and gaze predictions were fused using a Kalman Filter-based approach, and the positional error from using EMG-only was significantly higher than the fusion of EMG and gaze. The final target position Root Mean Squared Error (RMSE) decreased from 9.28 cm with an EMG-only prediction to 6.94 cm when using a gaze-EMG fusion. This error also increased significantly when removing some or all arm muscle signals. However, using fused EMG and gaze, there were no significant difference between predictors that included all muscles, or only a subset of muscles.**

Manuscript received December 11, 2019; revised April 13, 2020; accepted April 29, 2020. Date of publication May 6, 2020; date of current version June 5, 2020. This work was supported in part by the Swiss National Science Foundation through National Centre of Competence in Research in Robotics (NCCR Robotics) Wearable Scenario, NCCR Robotics Ph.D. Exchange Fellowship under Grant 51NF40–160592, and in part by the NIH T32 PRND Training Grant. *(Nili E. Krausz and Denys Lamotte are co-first authors.) (Corresponding author: Nili E. Krausz.)*

Nili E. Krausz was with the Shirley Ryan AbilityLab (formerly RIC), Chicago, IL 60611 USA, and also with the LASA Laboratory, EPFL, 1015 Lausanne, Switzerland. She is now with the Rehabilitation Institute of Chicago, Northwestern University, Chicago, IL 60611 USA, and also with the Weizmann Institute of Science, Rehovot 7610001, Israel (e-mail: nilikrausz2013@u.northwestern.edu).

Denys Lamotte, Iason Batzianoulis, and Aude Billard are with the LASA Laboratory, EPFL, 1015 Lausanne, Switzerland.

Levi J. Hargrove is with the Center for Bionic Medicine, Shirley Ryan AbilityLab (formerly RIC), Chicago, IL 60611 USA, and also with the Departments of Biomedical Engineering and Physical Medicine and Rehabilitation, Northwestern University, Chicago, IL 60611 USA.

Silvestro Micera is with the Center for Neuroprosthetics, EPFL, 1015 Lausanne, Switzerland, also with the Institute of Bioengineering, EPFL, 1015 Lausanne, Switzerland, and also with the Biorobotics Institute, Scuola Superiore Sant'Anna, 56025 Pontedera, Italy.

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2020.2992885

***Index Terms*— Prosthetics, upper limb prosthesis, electromyography, gaze tracking, sensory fusion, computer vision, end-point control, Kalman Filter.**

I. Introduction

ASSISTIVE technologies, such as prostheses [1] and exoskeletons [2], can help restore arm function and improve quality of life for individuals with amputations or other impairments. However, abandonment of these devices is high [3] largely due to poor controllability [4]. A robust, intuitive controller that reacts to user intention in a smooth, timely manner and provides natural motions for activities of daily living (ADLs), without placing added mental burden on the user, could improve quality of life for users.

We propose a novel full-arm prosthesis controller based on sensory fusion as shown in Figure 1. The overall objective is to predict the arm end-point and use a trajectory-based controller (such as inverse kinematics) to position individual arm joints. This could allow individuals with high level amputations to control an arm prosthesis with improved accuracy and speed without high cognitive burden.

This work focuses on characterizing gaze and EMG coordination among able-bodied participants during pick-and-place tasks, and evaluating an end-point prediction algorithm that could eventually be implemented in an arm prosthesis. In particular, we used a Kalman-based approach to fuse prediction of the end-effector position from a wearable head-mounted gaze tracking system and EMG sensors. We compared EMG-only predictions to those using an EMG-gaze fusion to evaluate the effects of introducing gaze to the controller.

II. Background

A. Myoelectric Control for Prosthetics

Electromyography (EMG) has been researched for control of prosthetic limbs for decades [5]–[8], with several approaches proposed for multifunctional control. One popular method has been dual-site proportional (or direct) control [9], in which EMG from an agonist-antagonist pair of muscles is mapped to the joint velocity of a single degree of freedom (DOF) [10]. A major limitation of this approach is the need for two independent muscular control sites for each prosthetic joint. This leads to a dimensionality mismatch, i.e., a higher number of desired outputs than the available

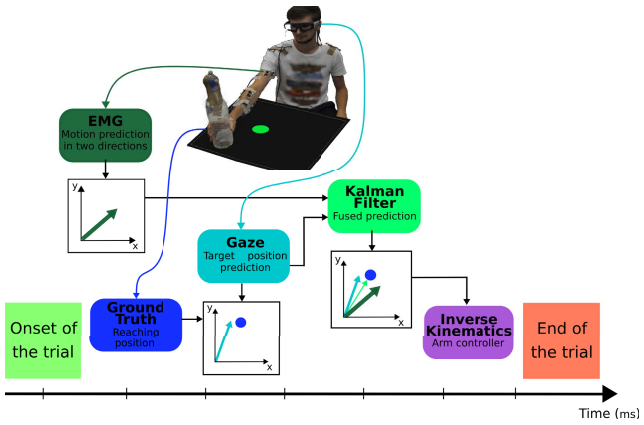

Fig. 1. Proposed end-point control strategy and experiment setup. In this system gaze and EMG are fused to estimate the position of a final target, and an inverse kinematics end-point controller is then used to drive the forward motion of the arm. Subjects were instrumented with 14 EMG channels on the shoulder, arm, and forearm, motion capture markers on the hand and shoulder for the ground truth estimation, and wearable gaze tracking glasses. Subjects followed a virtual target on a horizontal screen, and completed a total of 600 trials each. EMG is used to predict the direction of the motion and Gaze for predicting the target position. The predictions are fused with a Kalman Filter which produces a more certain prediction. The size of the arrows shows the uncertainty of each predictor.

independent inputs [4]. For a long time this has limited most powered prostheses to a single DOF. However, the increasing availability of multi-joint articulating hands [11]–[16] has driven research towards controlling additional DOFs.

Researchers have worked to address this limitation: 1) by increasing the number of usable muscle inputs, using novel surgical techniques [17], [18] or invasive interfaces [19], or 2) reducing output dimensionality, using principal component analysis (PCA) [20], [21], pattern classification techniques [22]–[25] (including the first commercially available system, COAPT [26]), or regression methods [27], [28]. Though the approaches in 1) can yield additional inputs, the quantity of these inputs remains limited, and although approaches listed in 2) can allow for control of multiple output DOFs but this is still limited to at most 3 DOFs at a time and users may have limited ability to independently control each input. Swain and Nightingale [29] proposed a system where orientation, trajectory, contact and other signals are combined with EMG to control 4-5 degrees of freedom, but this approach has yet to proven effective or clinically viable.

Full arm prostheses could potentially have as many as 6 DOFs proximal to the end-effector (shoulder rotation, humeral rotation, elbow, wrist flexion/extension, pronation/supination, and even wrist ulnar/radial deviation), and up to 19 DOFs [13] depending on the terminal device, so using a direct control or regression-based approach can easily eclipse the available control sites of most amputees. Thus, no clinically viable technology is available to allow an individual with a high-level amputation to control a full arm prosthesis with a multi-joint articulated hand. Importantly, there seems to be a high mental burden associated with trying to simultaneously direct the motion of multiple DOFs of a prosthesis.

B. Gaze Behavior During Pick-and-Place Tasks

Every day eye movements have been largely studied in [30]. During pick-and-place tasks gaze fixations are mainly associated with four roles: *locating* a desired object, *directing* the hand to an object/place, *guiding* two objects approaching each other and *verifying* the result of an action. Importantly, during the reaching, grasping, positioning or manipulation stages of any task the hands and objects in hand are never fixated on. Many daily tasks that seem to be autonomous, are in fact highly monitored by our vision. Thus, gaze tracking is a compelling option for intent prediction.

Using gaze as a control input is complicated by the so-called *Midas problem*, which refers to the difficulty of distinguishing a saccade involved in visual guidance from random or unrelated externally stimulated eye movements [31]. However, previous work has shown that fixations on objects of interest are usually much longer than fixations during manipulation tasks [32]. Furthermore, hand-eye coordination shows typical spatial patterns, indicating a correlation between eye motions and hand positioning. Indeed, the gaze path has a few important landmarks: the grasp site, the target (location where the object will be positioned) and any obstacles in the hand path. Importantly, gaze fixations on these landmarks are temporally correlated to hand motion, with a fixation always proceeding task execution, although anticipatory gaze behavior occurs during task observation as well [33].

Previously several works have proposed gaze-driven tracking for robotics contexts, such as in aiding human-robot interaction as in [34], [35]. Gaze-driven control has yet to be implemented in a control system for a powered prosthesis. Several previous works have considered using a fixed gaze tracking system for prosthetic or orthotic control based on a monitor mounted at some distance from the subject, including [36] and [37]. These have been promising; however, realistically for use with a prosthesis it is necessary for the entire system to ultimately be portable, and thus wearable head-mounted gaze systems are most promising for prosthetics. Additionally, requiring a stationary or restrained head is not clinically viable for a prosthesis control system.

However, head-mounted gaze behavior is complicated by head movements when not restrained. When the head is restrained the relationship between peak eye velocity and the corresponding eye motion amplitude during saccades is known, as is the relationship between duration and amplitude [38]. However, eye and head movements are controlled separately, thus no head kinematics can be deduced from eye motions in unrestrained conditions [39]. Additionally, it is suggested that head motions accompanying gaze movements appear regardless of motion amplitude [40]. Nevertheless, task execution is always regulated by the availability of the eyes which coordinate the task [41]. Thus, good gaze tracking performance requires a method for resolving head motion.

C. Multi-Modal Sensory Signal Fusion

Various human-generated signals have been considered for decoding user intent. Electro-encephalography (EEG) predicted the intention of reaching for an object almost 500 ms before movement onset [42]. Electro-corticographic signals (ECoG) allowed for identification of desired grasp type and the direction of reaching motion [43]. Similarly EMG

signals have been used to predict the desired grasp type prior to completion of the reaching motion in able-bodied users [44].

Thus, intention detection using a combination of sensory modalities remains less widespread despite potential advantages [45]. Combining EEG and EMG can lead to more stable intention detection performance compared to using a single modality alone [46]. Moreover, a combination of EEG recordings and eye tracking was used to drive an exoskeleton by triggering the motor activation using motor imagery detection but only tested with a few targets [47]. Nonetheless, studies typically use one modality as input to learning-based algorithms despite potential improvements in robustness, fatigue, or stability when using multiple modalities [48].

The human perceptual and motor systems are tightly coordinated [49], and this visuomotor coupling provides a potential control strategy for prosthetic devices. A combination of visual and motor signals might lead to more natural commands, as this reflects normal human control strategies and presumably the data sources would be extracted at the right time.

Fusion of multi-modal signals can be done in various ways. Raw data could be merged to train a machine learning algorithm with a new concatenated input [50] or features could first be extracted from each data source during a preprocessing step [51]. Nevertheless, a fusion of multi-modal data is complicated because data may not be available at exactly the same time, as is often the case when merging different physiological signals. In this case, the extracted features can first be separately classified and then the final decision can be made by combining classifier outputs [52]. This approach has shown particularly good results [53]. An advantage of using a classifier fusion is that the system remains robust even if one data source is not available or is noisy.

Previous work considered EMG and gaze as control inputs to a computer or virtual reality system, rather than a mouse or other input device, such as in [54], [55]. However, these works did not consider natural EMG and gaze behaviors present when completing a reach-to-grasp task, but rather require the user learn how to directly control several DoFs using EMG and gaze. This method maintains the control burden on the user, and does not leverage natural signals as intent cues to understand the desired goal of the reach-to-grasp task.

Recently, several groups have begun to consider combining gaze and EMG for prosthetic controls. Martin *et al.* [56] proposed the use of gaze and EMG to predict the end point position of a prosthesis; though in reality they located objects in the field of view as a means of approximating the gaze, this would be adversely affected by cluttered environments, and to our knowledge they did not quantify their system performance.

Meanwhile, Giordaniello, *et al.* [57] analyzed gaze and EMG during reach-to-grasp movements as a means of identifying the desired object in a group of objects for controlling a prosthesis in the MeganePro project; however, as discussed in Muller and Atzori [58], they did not present a means for fusing gaze and EMG, or results of using the two modalities for control. This work also only considered a traditional pattern-recognition-based EMG intent recognition approach, with 14 electrodes placed around the forearm, and thus was constrained to only predicting the desired hand grasp and not arm position.

III. Material and Methods

A. Experimental Protocol

Seven healthy subjects (4 males and 3 females, average age 26.8 ± 1.7) took part in the experiment. All were right-handed and used their dominant arm to perform the experiment. The Research Ethics Committee of EPFL approved the experiment and all subjects were consented prior to data collection. The data collection setup is showed in Figure 1.

Subjects sat at a table with a horizontal monitor and followed prompts from a custom-made user interface. Each trial started in a fixed resting position with the right palm placed on the screen. At the beginning of each trial block a plastic bottle was placed at the top center of the screen. A green circular target appeared and subjects were instructed to reach and grasp the bottle within 1200 ms. At this point a second green target appeared, and subjects had another 1200 ms to move the bottle and re-position it. Subjects were asked to maintain the position of the bottle for a short time until the target turned red, which was the prompt to return to the resting hand position. Targets appeared at random positions. To avoid fatigue, subjects were given a 4000 ms resting period before the next trial, which was cued when the target circle turned green. An experimental block was composed of 40 pick-and-place trials; and the total duration of each block of trials was 4 min 26 s. Each subject completed 12 experimental blocks and a total of 600 pick-and-place trials.

During the experiment, EMG signals from 14 upper limb muscles were recorded using a *DTS Noraxon system* [59], with a sampling rate of 1500 Hz. Muscles were chosen according to their role in arm motion (Supplementary Material, Table S1). We followed a standard procedure for surface EMG recordings: we preprocessed the data using a band-pass Butterworth filter between 20 Hz and 450 Hz to eliminate motion artifact, followed by full-wave rectification. A low-pass Butterworth filter with a 10 Hz cutoff frequency was applied to extract the signal envelope. Maximum voluntary contractions (MVC) were acquired for each muscle, and the preprocessed MVCs were then used to normalize each processed EMG channel.

Additionally, the subject's gaze was recorded with a wearable gaze tracker, the *SMI Eye Tracking Glasses* [60]. This device provides native, binocular tracking at up to 30 Hz, over the whole trackable field of view (80° horizontally and 60° vertically) with an accuracy of 0.5°. Two cameras inside the glasses frame record eye movements, while a forward-facing camera in the nose-piece records the field of view with a resolution of 1280*960 pixels. Included software automatically positions the wearer's gaze into the frame of the field of view, this position has been referred to as the Point of Regard (PoR) [61]. The gaze tracker was calibrated using a 3-point calibration method, where subjects were thrice instructed to fixate on a target while the experimenter selected the matching point in the streamed video of the wearer's field of view.

Hand position was recorded using an *Optitrack* [62] infrared motion capture system at 250 Hz. Two markers attached to the subject's hand and shoulder were used to record the relative

position of the hand with respect to shoulder. Hand position was used as ground truth for comparison in the analysis. An Arduino microcontroller was programmed to trigger the synchronization of the EMG recording and the motion capture system. The time stamp at the beginning of each trial was synchronized with the moment when the first target turned green, as visible in the video of the subject's field of view.

B. Gaze Event Detection

Natural gaze behavior involves a succession of fixations and saccades whose velocity profiles do not show any particular pattern during unconstrained head motions. Thus, neither the duration of a saccade nor its speed are useful for predicting movement. However, as has been shown, the fixations of a person's gaze appear to be closely related to the task [30]. This seems to be true both for points in the scene monitored by the subject and the relation between the fixation timing and task progress. (see section II-B) Therefore, to obtain valuable content from gaze tracking we extracted fixation events from raw gaze signals, which helps counter the *Midas problem*.

Detecting fixations from raw signals requires an algorithm that is robust to random eye movements and compensates for head motion. Three thresholds were defined, based on the assumption that points belonging to the same fixation are close to one another in the user's field of view.

1) *Maximum radius threshold*: The maximum radius that a fixation point cluster can have, with the mean defined as the center of the cluster. Points in the same fixation cluster can then be grouped. The radius threshold was set to 75 pixels, corresponding to a region of about 4% of the total image, which is about the size of object grasp sites, so only one fixation would be detected per object.
2) *Maximum step threshold*: This distinguishes points from two nearby fixations. As the head moves slower than the gaze, the distance between two points is greater if related to an eye saccade (50 pixels) than if due to a head motion (20 pixels). To distinguish these motions the maximum step threshold was set to 25 pixels.
3) *Duration threshold*: This removes fixations that are too rapid, as these are likely not related to the task. This was set to 300 ms given that the typical duration of a saccade can reach up to 250 ms.

The pseudo-code of the algorithm and an example of gaze event detection are respectively presented in the supplementary materials, Algorithm 1 and Figure S3.

C. Image Processing to Reduce Effect of Head Motion

Concurrent head motion during gaze displacement estimation can lead to under-estimation of the distance between two successive fixations in the user's field of view and overshadow the fixation detection algorithm. Therefore, we developed a computer vision algorithm to resolve concurrent head motion on the gaze tracking. We processed images from the front-facing camera to project the gaze position into a global 2D space (using the first recorded frame as a reference).

Visual features extracted from each image were matched to ones extracted from the previous image. Features were extracted with ORB [63], a binary descriptor based on the FAST corner detector [64] (see supplementary Figure S1) and the BRIEF descriptor [65], which was chosen for its rotation invariance and resistance to noise. This algorithm is computationally-efficient and can be used in real-time, which is important for future online implementation.

The FLANN algorithm [66], which constructs hierarchical k-means trees [67] to find an approximation of the nearest neighbors of each feature point, was then used to match features from consecutive frames. An example match is shown in supplementary materials Figure S2. RANSAC [68] was used to find the homographic transformation matrix to project each pixel from the head-frame to the global-frame, and produce gaze coordinates independent of head motions and solely dependent on the distance between targets in the global frame.

D. EMG Regression

Each preprocessed EMG channel was divided into sliding time windows of 200 ms, with an overlap of 100 ms and three time-domain features were extracted from each time window, as inputs to the regressor. Following previously described methods for EMG pattern recognition [69], the features chosen were waveform length, root mean square, and mean of each signal. With 14 channels and 3 features per channel, we obtained a 42-dimensional input for each time window. Inputs were scaled to a range of 0 to 1 for each dimension.

The mean hand position was computed for each window from the motion capture recordings. Rather than matching the EMG features to the position of the hand at the same time, the EMG time windows were matched to the hand position half a window later to account for the natural electromechanical delay estimated to be around 60-70 ms. [70] This allows for predicting hand position slightly in advance.

These inputs were provided to two ϵ-support vector regressors (ϵ-SVR) [71] trained for each subject, to predict the hand position in one of two dimensions (x and y). Four-fold cross-validation was used, with 75% of the data used for training and 25% for testing. The Radial Basis Function (RBF) kernel was used with each SVR to handle EMG non-linearities.

A grid search was performed for each subject to find the best hyperparameters of each SVR. The Mean Squared Error (MSE) computed on the testing set was used to evaluate each regressor, with lower errors corresponding better performance. These optimizations were done on the width of the RBF kernel γ, the soft margin parameter C and the ϵ parameter, which corresponds to the acceptable prediction error.

E. Gaze and EMG Fusion

Gaze fixations were fused with EMG-based predictions as described in Figure 1. The shift between two fixations in the subject's frame were matched to the amplitude of the hand motion in reference to the shoulder position. If a gaze fixation was detected, the direction of hand motion was updated by combining the EMG and Gaze predictions using a Kalman filter (see Figure 2). Every 100 ms we obtain a new hand motion direction estimate from EMG, and every 300 ms at minimum, we check for new gaze fixations.

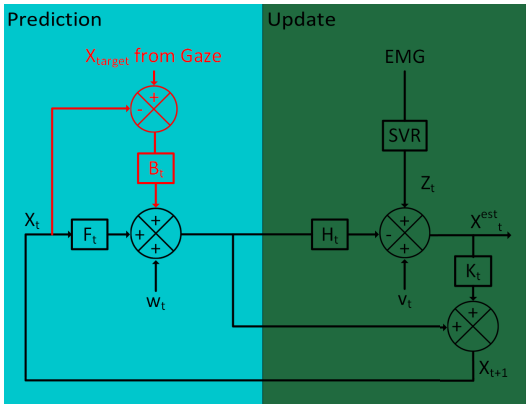

Fig. 2. Fusion algorithm. A Kalman Filter fused the gaze-tracking prediction (highlighted in red on left) with the SVR-based EMG prediction (in green box). The variances of the white noises (w_t and v_t) match the Mean Squared Error (MSE) of the gaze and EMG predictions. The measurement coefficient (H_t) is $[1; 1]$ because the predicted positions can directly be compared with the output of the SVRs, and the Kalman Gain (K_t) is computed at each step.

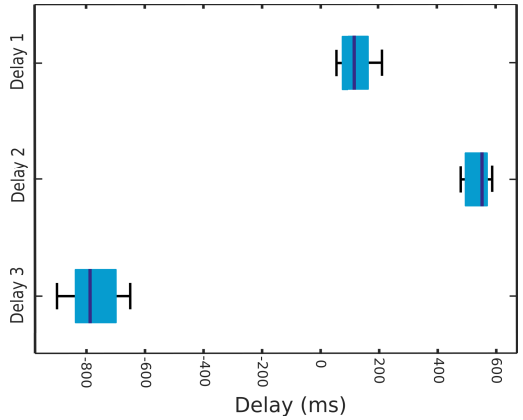

Fig. 3. Three Delays. Box-plots of the delays between the gaze fixation and the action for different steps of the motion. Delay 1 shows the delay between the end of the gaze fixation on the object and the end of the reaching motion. Delay 2 is between the beginning of the gaze fixation on the target and the end of the reaching motion. Delay 3 is between the onset of the gaze fixation on the target and the end of the placing motion.

The desired position is updated with the current observation based on the processed EMG signals (green box in Figure 2) every $\Delta_t = 100$ ms. The motion is initialized in the observed direction (x and y predictions) without addition of gaze estimation as follows: $X_t = X_{t-1} + \Delta_t * (\dot{X}_{t-1})$ (blue box of Figure 2). Once the gaze has fixated on the object, the gaze shift to the target can also be used to predict hand motion. The gaze-based prediction is introduced to predict the target position before the completion of the placement motion (red part of the Figure 2). The position prediction then follows $x_t = x_{t-1} + B_t * (x_{target} - x_{t-1})$, where B has been optimized based on the subject's residual error and x_{target} comes from the prediction made with gaze. The predicted position is then updated with the current EMG observation to produce an optimal fused predictor with gaze and EMG.

F. Analysis

Several analyses were performed to assess the gaze and EMG fusion-based prediction for control of a robotic arm:

1) *Timing*: Timing between gaze fixations and the task was evaluated to determine if the detected gaze is indeed related to the pick-and-place task, and whether the final object placement prediction occurs earlier using the fused system, given that gaze changes precede motion.
2) *Fixation Detection*: Evaluation of how well the fixation algorithm distinguishes gaze information related to the task (fixations on the object and target).
3) *Gaze and Motion Relationship*: The distance between two fixations in the subject's field of view, in terms of pixels, was analyzed to determine correlation with the motion amplitude during object placement.
4) *EMG Regression Performance*: Evaluation of the regression-based EMG predictors to determine how well the regression-based approach works for reaching tasks. These results could help refine the fusion algorithm.
5) *Fusion Performance*: The Mean Squared Error (MSE) provides the accuracy of the Kalman Filter-based position estimation, with lower values signifying better estimation. The MSE was selected as it incorporates the estimator variance and bias. The Root Mean Squared Error (RMSE) was also computed, as this providing accuracies in the same unit as the estimation. Performance was evaluated during the trial and at the final position.
6) *Simulated Amputation*: A "simulated amputation" analysis was performed, in which EMG from only shoulder muscles (S), or shoulder and upper arm muscles (S+A), were used to train predictors, with performance compared to a predictor including all muscles. This indicates the viability of this approach for high-level amputees.

IV. Results

Results from the analyses discussed in the Methods section (see section III-F) are included here.

A. Timing

Synchronization between gaze and motion was considered at three different moments in the pick-and-place task:

1) The delay between the end of the gaze fixation on the object and the end of the reaching motion.
2) The delay between the beginning of the gaze fixation on the target and the end of the reaching motion.
3) The delay between the onset of the gaze fixation on the target and the end of the placing motion.

For each of these assessments, a one-tailed t-test was performed to validate the hypothesis that the first two delays are in mean greater than zero, i.e. the end of the gaze fixation on the object and the beginning of the gaze fixation on the target precede the end of the reaching motion, and that the third one is lower than zero, i.e. the gaze precedes the placing motion. Figure 3 shows results for each subject.

(1) The average delay between the end of the gaze fixation on the object and the end of the reaching motion is 121 ± 58 ms, with $p < 0.01$. A positive delay here means that the gaze remains fixed on the object until the hand reaches it, confirming the synchronization we expected.

(2) The average delay between the onset of the gaze fixation on the target and the end of the reaching motion over all subjects was 535 ± 43 ms. For each subject $p \ll 0.01$. This positive delay shows that, as expected, the gaze begins fixating on the target after completion of the reaching motion.

(3) The average delay between the gaze fixation initiation on the target and the end of the placing motion is -780 ± 90 ms. This was significant for all subjects. Furthermore, we performed the location test for a mean of -300 ms, corresponding to the minimum duration for detecting a gaze fixation. The p-value for this test was significantly lower than 1% ($\simeq 10^{-50}$) for each subject. The means are also coherent across subjects except for subject #4, the subject with the shortest duration placement (486 ms against 606 ms on average).

Results showed that gaze remains fixed on the object until grasped and then moves to the target position prior to repositioning. It is important to note the distinction between delay 1 and 3, which are associated with the initial positioning and repositioning to the target, respectively. These are clearly two separate elements of reach-to-grasp tasks that must be distinguished for effective use of gaze tracking.

B. Fixation Detection

As discussed in the previous section, the algorithm used to detect gaze fixations should only identify events that are related to the task. To validate this hypothesis, fixations were considered for each trial. Given the task performed, the following fixations were expected for each trial:

1) a fixation on the object that will be grasped
2) a fixation on the target position to place the object
3) no fixations on any other point

Since users could potentially fixate on another location prior to moving the object, it is possible that a fixation ending before the subject grasps the object could be detected. Detection of such a fixation was considered correct since it corresponds to a gaze fixation before the onset of the reaching task; however, this type of fixation would not be used in an online control system, since no motion would be detected from EMG.

The points of gaze fixation given by our algorithm were compared to the labeled positions in the videos we recorded and annotated. Gaze fixations related to the task were correctly detected on average $85.36\% \pm 7.5\%$ of the time (see supplementary Table S1 for per-subject breakdown). Moreover, all fixations fit one of the three expected categories (see supplementary Table S2 summarizing these specific results). We detected a correct fixation on the object in $86.54\% \pm 7.18\%$ of the trials, and on the target in $84.18\% \pm 6.08\%$ of the trials. The difference between object and target fixation accuracies may be due to subjects not feeling the need to focus on the target location to complete the task accurately.

Additionally, for $80.5\% \pm 10.6\%$ of cases in which we detect only a single fixation, the distance between the initial and target positions was less than 5 cm. Since this distance was small, the steps of the gaze saccades were shorter than the maximum step threshold of our algorithm, thus we only detected one long fixation. This is a limitation of our method for very small displacements, which EMG signals alone also cannot accureately predict. These fixations were not considered belonging to either the object or the target.

Cases where no fixation was detected correspond to $0.77\% \pm 1.5\%$ of all the trials. This occurred for two subjects where the gaze was not fixated for a period of time of more than 300 ms during a trial. Visual inspection of these videos revealed that subjects were distracted by something during the experiment and were not focused on completing the task.

C. Gaze and Motion Relationship

The relationship between gaze shift (distance in pixels between gaze fixation on the object and the target) and the distance traveled by the hand during object placement was characterized. As it is difficult to extract information about depth using a single camera, we only considered the relationship between gaze and distance in the x-axis. This relationship was modeled using a linear regression, with $R^2 = 0.92 \pm 0.03$. Across subjects, the average RMSE was 7.2 ± 0.8 cm. Including gaze also gave a prediction of movement direction along the x-axis (left or right) with an error of approximately 5.3% across all trials and subjects.

D. EMG Regression Performance

The hyper-parameter $\epsilon = 3$ was defined to ensure sufficient precision. To obtain the optimal value of the RBF kernel hyper-parameter γ and the soft margin hyper-parameter C a grid search was performed for each subject and hyper-parameters were obtained which produced a small MSE along with a small standard deviation on the testing sets. The optimal values were close across all the subjects with $\gamma \in [1, 2]$ and $C \in [15, 25]$.

The testing and training set errors for all subjects are:

- x: RMSE $= 7.75 \pm 0.34$ cm (Test), 5.51 ± 0.07 cm (Train)
- y: RMSE $= 4.03 \pm 0.24$ cm (Test), 2.58 ± 0.05 cm (Train)

Thus, the error was almost twice as large for the x-axis as for the y-axis, for both testing and training sets. Though this could be due to the larger predictive range for the x-axis, when values were normalized by the width of the prediction range, the y-axis scored 0.082 and the x-axis scored 0.045.

Furthermore, the results across subjects are similar for training and testing sets, except for one subject for whom the MSE was much lower. Performing a one-way ANOVA on the data without this outlier subject gave a $p - value = 0.992$, supporting the hypothesis of equal MSEs between subjects.

E. Fusion Performance

Gaze predictions were fused with the EMG-based SVR predictions using the Kalman filter scheme described previously (see section III-E). First individual trajectories were considered, with two examples of ground truth and predicted trajectories, both with and without gaze, shown in Supplementary Material Figure S4. Differences between the EMG-only and the fused predictors were then considered across all trials, and results (shown in Figure 4) were compared to the motion capture-based ground truth using two different scores: the global MSE throughout the entire trial and the MSE computed at the target placement. In this figure it is clear that the global

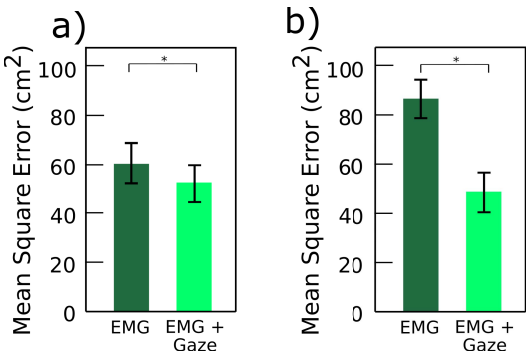

Fig. 4. Error for EMG alone or a Fusion of EMG and Gaze, shown for a) the entire motion, b) the final target position. For both a and b using EMG-only was significantly different from using the fusion of EMG and Gaze, for all subjects except for 6.

performance (Fig. S4a) improved with the addition of gaze as did the endpoint position accuracy (Fig S4b).

The global prediction of the motion along the x-axis across all subjects and trials decreased from an MSE $= 60.37 \pm 8.47$ cm^2 with the EMG-only prediction to MSE $= 52.14 \pm 7.72$ cm^2 with the fused gaze and EMG prediction. This was significant based on a one-sided t-test across trials (p-value < 0.05 for all the subjects except subject 6 p-value $= 0.09$) to test a decrease of 8 points. Thus, gaze improved continuous prediction of the placement motion. As seen in Figure S4a, movement to the target can begin once the gaze prediction has been made.

Additionally, we considered how fixation detection errors might have impacted the results. Examples where the target fixation incorrectly preceded the movement, leading to premature reaching motion predictions, were removed from the dataset. Removing these examples yielded a decrease of 14 points and an MSE of 46.87 ± 5.82 cm^2.

The final placement position prediction was also considered, and in this case the improvement from adding gaze was highly significant. The MSE of all subjects decreased from 86.21 ± 7.92 cm^2 to 48.48 ± 8.0 cm^2, an average decrease of more than 40%. Furthermore, one-sided t-tests across trials indicated that this result is significant (p-value < 0.05) for all but subject 6 (p-value $= 0.2$) for a decrease of 30 points. Finally, the RMSE (6.94 ± 0.58 cm) for the fused predictor is smaller than the error when using either an EMG-only or gaze-only predictor. Thus, fusion of the two predictors provides a significantly better accuracy than either individual predictor.

F. Simulated Amputation

A simulated amputation analysis was performed, with final target position error based on three different muscle groups, with both the EMG-only prediction and fused EMG-gaze prediction (Figure 5). Results for all muscles were detailed in section IV-E. Results were also considered for predictors that only included shoulder muscles (S) and for an intermediate, shoulder and upper arm muscles (S+A) set. When only shoulder muscles were considered the average MSE for the final target prediction decreased from 131.8 ± 12.6 cm^2 when using the EMG-only prediction to 71.4 ± 19.0 cm^2 when using the gaze and EMG fusion. Similarly with a muscle set

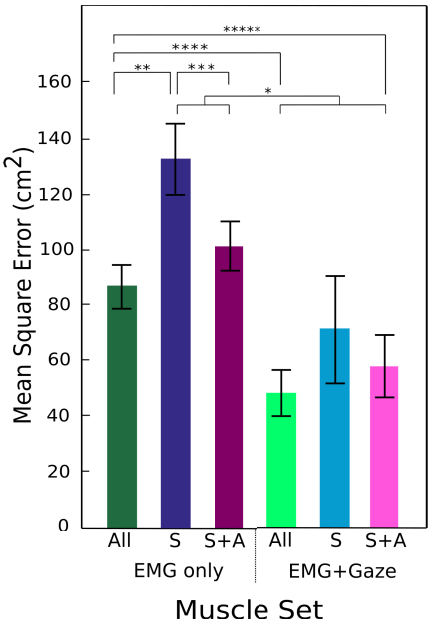

Fig. 5. Mean Squared Error for results of simulated amputation across different subjects. The averaged results across all subjects are shown with each muscle set labeled. The results for all muscles (labeled All) using EMG only is shown in dark green, and the results for all muscles using fusion of gaze and EMG is shown in light green. Results with only shoulder muscles (labeled S) using only EMG are shown in dark blue, and only shoulder muscles using a fusion of gaze and EMG is shown in light blue. Results with shoulder and arm muscles (labeled S+A) using only EMG are shown in dark pink, and shoulder and arm muscles using a fusion of gaze and EMG is shown in light pink. The addition of gaze information to the shoulder-only and the shoulder-and-arm condition has a significant impact on the MSE. Additionally, with EMG+gaze the three sets of muscles were no longer significantly different.

composed of muscles from the shoulder and arm, the average MSE decreased from 101 ± 8.9 cm^2 to 58 ± 11.1 cm^2 when including gaze data. Performing a t-test showed that the error decrease when including gaze is highly significant (p-value < 0.01) if we consider the same muscle set.

The results for each muscle set with EMG-only were also compared to those with a fusion of EMG and gaze. As expected, when fewer muscles are used the MSE increases. When only EMG was used in the prediction the results for all muscles and shoulder and arm muscles were significantly different from the MSE for shoulder muscles only. However, when a fusion of EMG and gaze was used there was no longer any significant difference between the three muscle groups (p-value > 0.65). Additionally, the error for the EMG-only muscle sets was significantly different than that of the EMG+Gaze sets for both All muscles and shoulder and arm muscles sets (p-value < 0.01), with lower errors reported for all muscle sets when incorporating gaze.

V. Discussion

The objective of this project was to understand coordination between gaze, EMG and arm motion during reaching and placement tasks and to develop a control system to seamlessly assist individuals with high level upper-limb amputations to use a full arm, multi-DOF prosthesis or orthosis. This work focused on predicting desired hand positions during pick-and-place tasks, using a fused gaze-EMG predictor. The analysis was offline, but could theoretically be performed online.

Our results indicate that gaze precedes hand motion during pick-and-place tasks. Gaze fixates on the desired object prior to the reaching motion, and then once the reaching motion occurs fixates on the successive target position, at which point re-positioning commences. The delay between the initial gaze fixation on the target and the end of the placing motion is sufficient for estimation of the hand motion. Using our algorithm, gaze fixations cannot be detected before 300 ms, however the average duration of the placement motion is 606 ms, so the delay between the gaze fixation and the placement motion is sufficient to estimate hand motion.

Across all subjects, gaze fixations were correctly detected in 86.6% $\pm$ 7.5% of trials using our algorithm, indicating good predictability, particularly given the high number of trials performed by each subject. However, a 14% error rate is still too high for a successful control system. Accuracy might improve by adding an object segmentation module that would locate objects or targets within the field of view, which in combination with gaze and EMG might allow for more precise and better estimates of when grasping is desired. Ultimately, the performance may also change substantially when operating online, but this will only become clear upon testing.

Additionally, the relationship between the distance between gaze fixations and the motion size were modeled using a linear regression, and this relationship was found to be quite linear ($R^2 = 0.92 \pm 0.03$) with a RMSE $= 7.2 \pm 0.8$ cm, thus the shift between gaze fixations could be used to estimate arm motion.

EMG was used to train two SVRs, and the performance of these predictors was also analyzed. The RMSE errors for the testing sets for the x-axis and y-axis were 7.75 ± 0.34 cm and 4.03 ± 0.24 cm, respectively; this disparity between x and y held true even when normalized by the range (which differed for x and y). Thus, the better estimation obtained for the y-axis must also be related to the complexity of the motion, which is important as the gaze is not useful for providing estimation in the y- direction. Thus the x-prediction was supplemented by gaze, but the y-prediction was not. Additionally, the results of the ANOVA to compare the regression between subjects was not significant, which suggests that this approach is subject-independent despite the diversity that existed between subjects in terms of size, gender, fitness, and other factors.

The regression testing set error was twice the training set error, with or without optimizing the SVR parameters. The general applicability of our model is constrained by the fact that for each target position there are multiple previous positions, leading to different movement paths, and correspondingly different EMG signals across trials. This variation in paths taken may have led to reduced predictor performance.

Finally, the results of the fusion algorithm have several meaningful implications; such as the addition of gaze reducing the prediction error throughout the trials, as well as in the final target position. More specifically, the average RMSE across all subjects decreased from approximately 9.28 cm to 6.94 cm for the final x-position, and the average RMSE was 4.03 cm for the final y-position. Thus, gaze information increased the target prediction accuracy compared to EMG-only.

In contrast, results during the trials did not show a large improvement, likely due to target fixations detected before the motion, leading to the incorrectly predicting reaching before any movement. This increased the error even if the final position predictions remained good. The RMSE for the dataset without these erroneous fixations was 6.85 ± 2.4 cm, significantly lower than the full dataset (7.8 ± 2.9 cm).

The importance of including the gaze is illustrated by the predicted trajectories as shown in the examples of the Results section. In cases where the EMG-only prediction diverged from the trajectory and the final target, the fused EMG-gaze prediction remained closer to the ground truth. Both the position and direction predictions were improved by the inclusion of gaze.

Another important result is that the overall error in each direction is less than 10 cm. Ideally the RMSE would be close to zero; however, our approach did provide a means to predict the desired end-point with relatively good accuracy using this approach. Further analysis will focus on improving the prediction accuracy, using Computer Vision (see Future Work VI). Additionally, other methods of fusing gaze and EMG data will be considered for minimizing errors.

This work was somewhat limited in that we only used able-bodied data, and some differences are likely when using a similar approach with individuals with disabilities, such as stroke or amputation. To address this concern, an analysis was performed to simulate how the regression would perform using a limited muscle set. Results showed that the errors when using only the shoulder muscles or only the shoulder and upper arm muscles were higher, however by using the gaze and EMG fusion this effect was mitigated. In fact, when using the fused predictor that included gaze and EMG, there was no significant difference between the three muscle sets compared.

Additionally, for each muscle set the fused predictor yielded an MSE significantly lower than the EMG-only predictor. The average RMSE increased from 6.94 cm to 7.6 cm to 8.3 cm respectively, for the all muscles, shoulder+arm muscles, and shoulder-only muscle groups when using the fusion predictor. The error only increased by < 3 cm when eliminating all arm muscles, indicating that gaze may compensate for reduced muscles availability of high level amputees.

However, the system may be affected by changes in EMG when used by amputee subjects, such as the added weight of wearing a prosthesis or lack of natural EMG and movement patterns that exist when performing a reaching motion with a full arm. Additionally, the lack of confidence many amputees have in their prostheses may lead to longer than normal fixations on the prosthesis. We believe that over time if a prosthesis performs well enough and intuitively then gaze fixations would likely become more similar to able-bodied timing. To this end it may be necessary to devise novel training methods to allow for reduced prosthesis fixations over time, possibly by using VR to simulate where the prosthetic hand would be, or by using visualization or mirror practice. Alternatively, the control system may require major adjustments for individuals who've become accustomed to fixating heavily on their prostheses, so additional testing will

be necessary to validate this approach with amputee subjects and individuals with other disabilities.

VI. Future Work

We aim to test this approach online with healthy subjects and an external robotic arm and hand. A shared controller will be used, with our algorithm predicting the desired end-point position, and an autonomous system (inverse kinematics-driven) controlling individual degrees of freedom to position the end-point in the desired location. A gaze-EMG fusion will signal the desired target; however, a bio-signal (such as EMG) may be required to trigger the motion (to prevent unintentional movements). For instance, the arm motion initiation might wait until EMG prediction surpasses a threshold. A similar approach has shown interesting results with a combination of EMG and a head mounted sensor used for hand preshaping control [72]. To our knowledge, this approach has not been used to control a reach-to-grasp task with a full arm.

Vision may be used to improve target accuracy. Object segmentation, if performed effectively could allow for more accurate estimation of the desired target location. Knowing the object and end effector positions could also allow for accurate completion of reaching. Moreover, object segmentation could be useful for grasp-type classification, and will be determined through online testing with able-bodied and amputee subjects.

Several metrics will be used to evaluate system performance for use in a real-time, commercial control system. First, target prediction accuracy will be compared to the motion capture ground truth, and online results will validate our offline estimates. Second, hand trajectories and movement smoothness will be analyzed. Third, task completion time will be evaluated, with characterization of the human and robot delays. Generally, a delay of less than 200 ms is considered acceptable for a responsive system. Finally, functional outcomes of the control and a conventional direct control approach will be compared using standard outcomes measures such as the Box and Blocks test, Clothespin test, Southampton Assessment Protocol (SHAP), and self-reported questionnaires.

VII. Conclusion

Current myoelectric systems are limited from controlling multi-DOF devices by the number of control signals available. We propose a novel shared-control strategy for an arm prosthesis, in which gaze and EMG data are used to estimate the desired target position for the end-point of an arm and a dynamical model is used to drive the forward motion of the arm. We analyzed gaze and EMG coordination during pick-and-place tasks, using a virtual interface and a motion capture system. EMG was used to train two SVRs to predict hand position in the x- and y-directions, and a Kalman Filter-based approach was used to fuse these estimates with a prediction, based on the relationship between gaze shifts and arm motion. Fusing gaze and EMG produced higher accuracy position estimates than using EMG alone, and simulations of high-level arm amputation resulted in a small decrease in accuracy, with the inclusion of gaze more than compensating for the reduced number of EMG sites, suggesting that this approach could be used clinically. Future work will focus on implementing this approach in an online system with a robotic arm, and testing will be completed with individuals with limb loss.

Acknowledgment

The authors would like to thank S. Varonov for his assistance with the gaze filtering algorithm, and A. Barlow for assisting with editing and revising this paper.

References

[1] O. Van Der Niet Otr, H. A. Reinders-Messelink, R. M. Bongers, H. Bouwsema, and C. K. Van Der Sluis, "The i-LIMB hand and the DMC plus hand compared: A case report," *Prosthetics Orthotics Int.*, vol. 34, no. 2, pp. 216–220, Jun. 2010.

[2] H. S. Lo and S. Q. Xie, "Exoskeleton robots for upper-limb rehabilitation: State of the art and future prospects," *Med. Eng. Phys.*, vol. 34, no. 3, pp. 261–268, Apr. 2012.

[3] K. A. Raichle, "Prosthesis use in persons with lower- and upper-limb amputation," *J. Rehabil. Res. Develop.*, vol. 45, no. 7, pp. 961–972, Dec. 2008.

[4] E. Biddiss, D. Beaton, and T. Chau, "Consumer design priorities for upper limb prosthetics," *Disab. Rehabil., Assistive Technol.*, vol. 2, no. 6, pp. 346–357, Jan. 2007.

[5] B. Hudgins, P. Parker, and R. Scott, "Control of artificial limbs using myoelectric pattern recognition," *Med. Life Sci. Eng.*, vol. 13, pp. 21–38, Jan. 1994.

[6] M. Rakić, "Report on the further development of the belgrade hand prosthesis," in *Proc. Int. Symp. External Control Hum. Extremities, Dubrovnik*, 1966, pp. 90–96.

[7] P. Herberts, C. Almström, R. Kadefors, and P. D. Lawrence, "Hand prosthesis control via myoelectric patterns," *Acta Orthopaedica Scandinavica*, vol. 44, nos. 4–5, pp. 389–409, Jan. 1973.

[8] J. M. Nightingale, "Microprocessor control of an artificial arm," *J. Microcomput. Appl.*, vol. 8, no. 2, pp. 167–173, Apr. 1985.

[9] L. H. Smith, T. A. Kuiken, and L. J. Hargrove, "Real-time simultaneous and proportional myoelectric control using intramuscular EMG," *J. Neural Eng.*, vol. 11, no. 6, Dec. 2014, Art. no. 066013.

[10] P. Parker, K. Englehart, and B. Hudgins, "Myoelectric signal processing for control of powered limb prostheses," *J. Electromyogr. Kinesiol.*, vol. 16, no. 6, pp. 541–548, Dec. 2006.

[11] RSLSteeper. (2017). *Bebionic Kernel Description*. [Online]. Available: http://bebionic.com/

[12] T. Bionics. (2017). *iLimb Ultra Kernel Description*. [Online]. Available: http://www.touchbionics.com/products/active-prostheses/i-limb-ultra

[13] R. Weir *et al.*, "The intrinsic hand—A 22 degree-of-freedom artificial hand-wrist replacement," in *Proc. Myoelectr. Symp.*, 2008, pp. 1–5.

[14] L. Resnik, S. L. Klinger, and K. Etter, "The DEKA arm: Its features, functionality, and evolution during the veterans affairs study to optimize the DEKA arm," *Prosthetics Orthotics Int.*, vol. 38, no. 6, pp. 492–504, Dec. 2014.

[15] M. Controzzi, F. Clemente, D. Barone, A. Ghionzoli, and C. Cipriani, "The SSSA-MyHand: A dexterous lightweight myoelectric hand prosthesis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 5, pp. 459–468, May 2017.

[16] J. T. Belter, J. L. Segil, A. M. Dollar, and R. F. Weir, "Mechanical design and performance specifications of anthropomorphic prosthetic hands: A review," *J. Rehabil. Res. Develop.*, vol. 50, no. 5, p. 599, 2013.

[17] T. A. Kuiken, "Targeted muscle reinnervation for real-time myoelectric control of multifunction artificial arms," *JAMA*, vol. 301, no. 6, pp. 619–628, Feb. 2009.

[18] M. G. Urbanchek, B. Wei, Z. Baghmanli, K. Sugg, and P. S. Cederna, "Long-term stability of regenerative peripheral nerve interfaces (RPNI)," *Plastic Reconstructive Surg.*, vol. 128, no. 4S, pp. 88–89, Oct. 2011.

[19] R. F. Weir, P. R. Troyk, G. DeMichele, and D. Kerns, "Technical details of the implantable myoelectric sensor (IMES) system for multifunction prosthesis control," in *Proc. IEEE Eng. Med. Biol. 27th Annu. Conf.*, Jan. 2005, pp. 7337–7340.

[20] J. L. Segil, S. A. Huddle, and R. F. F. Weir, "Functional assessment of a myoelectric postural controller and multi-functional prosthetic hand by persons with trans-radial limb loss," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 618–627, Jun. 2017.

[21] G. C. Matrone, C. Cipriani, E. L. Secco, G. Magenes, and M. C. Carrozza, "Principal components analysis based control of a multi-dof underactuated prosthetic hand," *J. NeuroEng. Rehabil.*, vol. 7, no. 1, p. 16, Dec. 2010.

[22] M. A. Oskoei and H. Hu, "Myoelectric control systems—A survey," *Biomed. Signal Process. Control*, vol. 2, no. 4, pp. 275–294, 2007.

[23] A. B. Ajiboye and R. F. H. Weir, "A heuristic fuzzy logic approach to EMG pattern recognition for multifunctional prosthesis control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no. 3, pp. 280–291, Sep. 2005.

[24] A. Hiraiwa, K. Shimohara, and Y. Tokunaga, "EMG pattern analysis and classification by neural network," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Nov. 1989, pp. 1113–1115.

[25] A. J. Young, L. H. Smith, E. J. Rouse, and L. J. Hargrove, "Classification of simultaneous movements using surface EMG pattern recognition," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 5, pp. 1250–1258, May 2013.

[26] Coapt Engineering. (2017). *COAPT Kernel Description*. [Online]. Available: https://coaptengineering.com/

[27] S. Amsuess *et al.*, "Context-dependent upper limb prosthesis control for natural and robust use," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 7, pp. 744–753, Jul. 2016.

[28] L. H. Smith, T. A. Kuiken, and L. J. Hargrove, "Evaluation of linear regression simultaneous myoelectric control using intramuscular EMG," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 737–746, Apr. 2016.

[29] I. D. Swain and J. M. Nightingale, "An adaptive control system for a complete hand/arm prosthesis," *J. Biomed. Eng.*, vol. 2, no. 3, pp. 163–166, Jul. 1980.

[30] M. Land, N. Mennie, and J. Rusted, "The roles of vision and eye movements in the control of activities of daily living," *Perception*, vol. 28, no. 11, pp. 1311–1328, Nov. 1999.

[31] R. Radach, J. Hyona, and H. Deubel, *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*. Amsterdam, The Netherlands: Elsevier, 2003.

[32] R. S. Johansson, G. Westling, A. Bäckström, and J. R. Flanagan, "Eye–hand coordination in object manipulation," *J. Neurosci.*, vol. 21, no. 17, pp. 6917–6932, 2001.

[33] B. Gesierich, A. Bruzzo, G. Ottoboni, and L. Finos, "Human gaze behaviour during action execution and observation," *Acta Psychol.*, vol. 128, no. 2, pp. 324–330, Jun. 2008.

[34] D. P. Noonan, G. P. Mylonas, A. Darzi, and G.-Z. Yang, "Gaze contingent articulated robot control for robot assisted minimally invasive surgery," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2008, pp. 1186–1191.

[35] R. Atienza and A. Zelinsky, "Active gaze tracking for human-robot interaction," in *Proc. 4th IEEE Int. Conf. Multimodal Interfaces*, Oct. 2002, pp. 261–266.

[36] E. A. Corbett, K. P. Kording, and E. J. Perreault, "Real-time evaluation of a noninvasive neuroprosthetic interface for control of reach," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 4, pp. 674–683, Jul. 2013.

[37] C. Castellini, "Gaze tracking in semi-autonomous grasping," *J. Eye Movement Res.*, vol. 2, no. 4, pp. 1–7, Nov. 2008.

[38] J. Van Gisbergen, J. Van Opstal, and F. Ottes, "Parametrization of saccadic velocity profiles in man," *Adv. Psychol.*, vol. 22, pp. 87–94, Jan. 1984.

[39] E. G. Freedman and D. L. Sparks, "Coordination of the eyes and head: Movement kinematics," *Exp. Brain Res.*, vol. 131, no. 1, pp. 22–32, Mar. 2000.

[40] J. Pelz, M. Hayhoe, and R. Loeber, "The coordination of eye, head, and hand movements in a natural task," *Exp. Brain Res.*, vol. 139, no. 3, pp. 266–277, Aug. 2001.

[41] J. Epelboim *et al.*, "The function of visual search and memory in sequential looking tasks," *Vis. Res.*, vol. 35, nos. 23–24, pp. 3401–3422, Dec. 1995.

[42] E. Lew, R. Chavarriaga, S. Silvoni, and J. D. R. Millán, "Detection of self-paced reaching movement intention from eeg signals," *Frontiers Neuroeng.*, vol. 5, p. 13, Jul. 2012.

[43] T. Pistohl, A. Schulze-Bonhage, A. Aertsen, C. Mehring, and T. Ball, "Decoding natural grasp types from human ECoG," *NeuroImage*, vol. 59, no. 1, pp. 248–260, Jan. 2012.

[44] I. Batzianoulis, S. El-Khoury, E. Pirondini, M. Coscia, S. Micera, and A. Billard, "EMG-based decoding of grasp gestures in reaching-to-grasping motions," *Robot. Auton. Syst.*, vol. 91, pp. 59–70, May 2017.

[45] N. Jiang, S. Dosen, K.-R. Müller, and D. Farina, "Myoelectric control of artificial limbs—Is there a need to change focus," *IEEE Signal Process. Mag*, vol. 29, no. 5, pp. 150–152, Sep. 2012.

[46] R. Leeb, H. Sagha, R. Chavarriaga, and J. R. del Millán, "A hybrid brain–computer interface based on the fusion of electroencephalographic and electromyographic activities," *J. Neural Eng.*, vol. 8, no. 2, 2011, Art. no. 025011.

[47] A. Frisoli *et al.*, "A new gaze-BCI-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 6, pp. 1169–1179, Nov. 2012.

[48] D. Novak and R. Riener, "A survey of sensor fusion methods in wearable robotics," *Robot. Auton. Syst.*, vol. 73, pp. 155–170, Nov. 2015.

[49] M. M. Hayhoe, A. Shrivastava, R. Mruczek, and J. B. Pelz, "Visual memory and motor planning in a natural task," *J. Vis.*, vol. 3, no. 1, p. 6, Feb. 2003.

[50] Y. Wu, E. Blasch, G. Chen, L. Bai, and H. Ling, "Multiple source data fusion via sparse representation for robust visual tracking," in *Proc. 14th Int. Conf. Inf. Fusion (FUSION)*, 2011, pp. 1–8.

[51] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proc. IEEE*, vol. 85, no. 1, pp. 6–23, Jan. 1997.

[52] D. Ruta and B. Gabrys, "An overview of classifier fusion methods," *Comput. Inf. Syst.*, vol. 7, no. 1, pp. 1–10, 2000.

[53] E. Erzin, Y. Yemez, and A. M. Tekalp, "Multimodal speaker identification using an adaptive classifier cascade based on modality reliability," *IEEE Trans. Multimedia*, vol. 7, no. 5, pp. 840–852, Oct. 2005.

[54] Y. S. Pai, T. Dingler, and K. Kunze, "Assessing hands-free interactions for VR using eye gaze and electromyography," *Virtual Reality*, vol. 23, no. 2, pp. 119–131, Jun. 2019.

[55] J. C. Mateo, J. S. Agustin, and J. P. Hansen, "Gaze beats mouse: Hands-free selection by combining gaze and EMG," in *Proc. Extended Abstr. Hum. Factors Comput. Syst.*, 2008, pp. 3039–3044.

[56] H. Martin, J. Donaw, R. Kelly, Y. Jung, and J.-H. Kim, "A novel approach of prosthetic arm control using computer vision, biosignals, and motion capture," in *Proc. IEEE Symp. Comput. Intell. Robot. Rehabil. Assistive Technol. (CIRAT)*, Dec. 2014, pp. 26–30.

[57] F. Giordaniello *et al.*, "Megane pro: Myo-electricity, visual and gaze tracking data acquisitions to improve hand prosthetics," in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Jul. 2017, pp. 1148–1153.

[58] H. Müller and M. Atzori, "Combining object recognition, gaze tracking and electromyography to guide prosthetic hands–experiences from two research projects," *Res. Hybrid Control Neuro-Prosthetics*, vol. 7, p. 15, May 2019.

[59] *Surface EMG*. Accessed: Apr. 30, 2020. [Online]. Available: http://www.noraxon.com/

[60] *Tobii Pro Glasses 2*. Accessed: Apr. 30, 2020. [Online]. Available: https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/

[61] M. Sobuh, L. Kenney, A. Galpin, S. Thies, P. Kyberd, and R. Raffi, "Coding scheme for characterising gaze behaviour of prosthetic use," in *Proc. Myoelectr. Controls/Powered Prosthetics Symp. Fredericton*, Fredericton, NB, Canada, Aug. 2011, pp. 14–19.

[62] *OptiTrack*. Accessed: Apr. 30, 2020. [Online]. Available: http://www.optitrack.com/

[63] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2564–2571.

[64] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2006, pp. 430–443.

[65] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2010, pp. 778–792.

[66] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," *VISAPP*, vol. 2, nos. 331–340, p. 2, 2009.

[67] K. Fukunaga and P. M. Narendra, "A branch and bound algorithm for computing k-Nearest neighbors," *IEEE Trans. Comput.*, vol. C-24, no. 7, pp. 750–753, Jul. 1975.

[68] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[69] M. Ahsan, M. I. Ibrahimy, and O. O. Khalifa, "EMG signal classification for human computer interaction: A review," *Eur. J. Sci. Res.*, vol. 33, no. 3, pp. 480–501, 2009.

[70] P. R. Cavanagh and P. V. Komi, "Electromechanical delay in human skeletal muscle under concentric and eccentric contractions," *Eur. J. Appl. Physiol. Occupational Physiol.*, vol. 42, no. 3, pp. 159–163, Nov. 1979.

[71] A. Smola and V. Vapnik, "Support vector regression machines," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 9, 1997, pp. 155–161.

[72] C. Cipriani, F. Zaccone, S. Micera, and M. C. Carrozza, "On the shared control of an EMG-controlled prosthetic hand: Analysis of user–prosthesis interaction," *IEEE Trans. Robot.*, vol. 24, no. 1, pp. 170–184, Feb. 2008.