Effectiveness of different sensing modalities in predicting targets of reaching movements*

Domen Novak, Ximena Omlin, Rebecca Leins-Hess, and Robert Riener, *Member, IEEE*

***Abstract*— Human motion recognition is essential for many biomedical applications, but few studies compare the abilities of multiple sensing modalities. This paper thus evaluates the effectiveness of different modalities when predicting targets of human reaching movements. Electroencephalography, electrooculography, camera-based eye tracking, electromyography, hand tracking and the user's preferences are used to make predictions at different points in time. Prediction accuracies are calculated based on data from 10 subjects in within-subject crossvalidation. Results show that electroencephalography can make predictions before limb motion onset, but its accuracy decreases as the number of potential targets increases. Electromyography and hand tracking give high accuracy, but only after motion onset. Eye tracking is robust and gives high accuracy at limb motion onset. Combining multiple modalities can increase accuracy, though not always. While many studies have evaluated individual sensing modalities, this study provides quantitative data on many modalities at different points of time in a single setting. The information could help biomedical engineers choose the most appropriate equipment for a particular application.**

Keywords—intention detection, reaching movements, eye tracking, electroencephalography, electromyography

I. INTRODUCTION

Recognition of voluntary movements is crucial in many fields of biomedical engineering, and many different technologies have been used for this purpose. For example, neuroprostheses can be controlled with brain-machine interfaces [1] while exoskeletons use electromyography to augment movement [2]. In each application, it is necessary to carefully select the most appropriate sensing technology. Selection criteria include the cost of the equipment, the user-friendliness, the accuracy and the time at which the movement can be recognized (before/after movement onset).

One important type of voluntary movements is reaching movement toward one of several objects. Faced with many possible targets, the goal is to predict the actual target quickly and accurately. This information can then be used by e.g. assistive devices that support the movement. Many modalities can be used for prediction. When choosing among objects, a person's gaze shifts toward the eventually chosen item [3–4], which can be measured using eye trackers. Brain activity associated with motion planning can be measured with electroencephalography [5–7]. Electrical muscle activity precedes limb motion onset [2, 8], and movement direction indicates the target before it is reached.

All these possibilities have been extensively studied, and attempts have been made to combine multiple modalities. For example, Corbett et al. [9] used eye tracking to predict the reaching target and then planned the movement trajectory using electromyography. However, no studies directly compare numerous sensing modalities in a single setting with regard to two important criteria: how accurate they are and how early they can predict the target of the movement.

Our study attempts to address this issue by predicting targets of reaching movements using electroencephalography (EEG), electrooculography (EOG), camera-based eye tracking, electromyography (EMG), hand tracking and the user's preferences. Each modality is evaluated at multiple points in time individually and in combination with others.

II. MATERIALS AND METHODS

A. Subjects

Ten healthy right-handed subjects (9 males, 1 female, 27.0 ± 2.4 years old) participated in the study.

B. Task environment

The task consisted of multiple reaching trials performed in a horizontal plane using virtual objects. The subject sat at a desk with the right hand on the desktop (Fig. 1). A screen on the desk displayed a virtual environment (Fig. 2) that initially consists of 3 platters and a red square that marks the starting position for each movement. Sixteen possible objects representing reaching targets can appear on the platters. They are divided into 7 categories: fruit, meat, nonalcoholic drinks, alcoholic drinks, sweets, toys, and cigarettes.

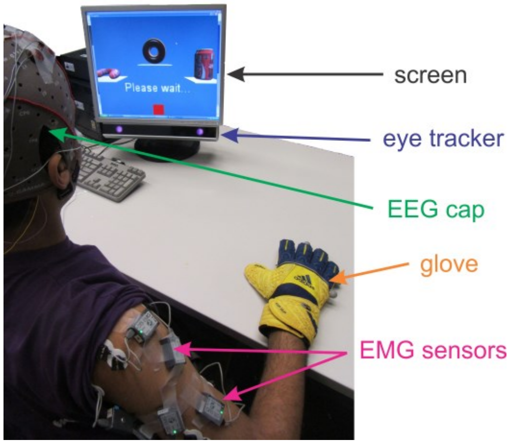

Figure 1. Experiment setup.

* This work was supported by the Swiss National Science Foundation through the National Centre of Competence in Research Robotics.

All authors are with ETH Zurich, Switzerland (phone: +41 7744-70158; e-mail: firstname.lastname@hest.ethz.ch).

Figure 2. Virtual task environment, with examples of objects.

The subject's hand was tracked and shown in the virtual environment as a red pointer. Left/right hand movement moves the pointer horizontally while movement away/toward the subject moves the pointer vertically.

C. Signal acquisition

Two *g.USBamp* systems (g.tec Medical Engineering, Austria) recorded EEG and EOG at 600 Hz. EEG was measured with the g.GAMMAcap and g.Butterfly electrodes. Electrodes were placed at locations F1, F2, F3, F4, Fz, FC1, FC4, C1, C3, CP1, CP3, CPz, P1, P3, P4 and PO3 of the 10-20 system since information about movement direction can be found in those areas [5–7]. The reference was placed at Cz. EOG was measured with electrodes to the upper right and lower left of the eyes. EEG was bandpass-filtered from 0.5 to 40 Hz. EOG was lowpass-filtered with a cutoff at 40 Hz.

The *SMI RED* (SensoMotoric Instruments, Germany), a remote eye tracking system, was mounted underneath the screen and tracked gaze position at a frequency of 60 Hz.

The *Telemyo 2400 DTS* (Noraxon, USA) recorded EMG at 1.5 kHz. Electrodes were placed on the anterior, posterior and medial deltoid, pectoralis major, infraspinatus, biceps brachii, lateral and medial triceps using SENIAM placement [10]. EMG was bandpass-filtered from 10 to 500 Hz [10].

The *QualiSys Oqus* (Qualisys AB, Sweden), a system of infrared cameras and passive markers, tracked hand position at a frequency of 60 Hz. Cameras were placed in the room, and a glove with rigid markers was placed on the hand.

D. Questionnaire

Subjects were given a 9-item questionnaire. The first two questions were "How hungry/thirsty are you right now?" while the others were "How much do you like alcohol / cigarettes / meat / fruit / sweets / toys / nonalcoholic drinks?" Answers were given on a 5-point scale, with a special score of 0 if the subject does not consume that type of object at all. Each virtual object was then assigned three values: a preference, a need, and a product of the need and preference. The preference was equal to the relevant "How much do you like..." answer. Need was equal to the relevant hunger/thirst value: hunger for fruit, meat and sweets and thirst for drinks. Cigarettes were assigned a need of 3 for smokers and 0 otherwise. Toys were always assigned a need of 3.

E. Measurement protocol

Upon the subject's arrival, the purpose and procedure of the study were explained. The hardware was applied and the questionnaire was given. Subjects then performed 40 trials: 20 trials with 2 visible objects ("2-object trials") and 20 trials with 3 visible objects ("3-object trials"). Subjects were instructed to reach for the object they would prefer to have. Displayed objects were always from different categories.

In each trial, the subject first rested for 5 s. The objects then appeared on the screen. The subject was shown a "please wait" message below the objects and mentally chose an object without moving. After 5 s, the message disappeared and the subject reached for the chosen object. Upon reaching it, the subject returned to the starting point.

F. Target prediction

Target prediction rules were obtained from recorded training data using supervised machine learning. Features were extracted from raw signals and input to a classifier. Prediction accuracy was obtained with crossvalidation.

F.1 Feature extraction

Features represent relevant information extracted from raw data. Most of these features are calculated over a window of time up to the present - the interval [t-L, t]. Multiple values of L (100-1000 ms) were tested in crossvalidation (section F.3).

EEG was cleaned of eye artifacts using an adaptive filter with EOG as the noise input [11]. Three feature types were then extracted for each channel over a window*: root-mean-square values*, *mean frequencies* (with Welch's method), and *autoregressive coefficients* (third-order, Burg method [6]).

EOG consists of horizontal and vertical components. The *mean values* of both were calculated over a window.

Eye tracker software outputs *gaze position on the screen*. Its mean value was calculated over a window.

EMG features include each channel's *root-mean-square value* and *variance* calculated over a window.

Hand tracking features include the *current hand position* and *current hand velocity* in x and y coordinates.

Questionnaire features include the *differences in preference, need, and need-preference products* between objects 1 and 2 and between objects 1 and 3 in each trial.

F.2 Classification

Classification consists of two parts. From the entire feature set, relevant features are first selected using Sequential Floating Forward Selection (SFFS). This feature subset is then input into a linear discriminant analysis (LDA) classifier that outputs the predicted target. Both SFFS and LDA are commonly used with physiological data [12, 13].

Classifiers were trained for each individual sensing modality and for a few modality combinations at 7 possible points in time. These *combinations* are: questionnaire + EEG; questionnaire + hand tracking; EEG + EOG; EEG +

EMG; eye tracking + EMG; eye tracking + hand tracking; EMG + hand tracking; all data.

The 7 *points in time* are: 1 s / 3 s / 5 s after objects appear; at motion onset; after 25% / 50 % / 75% of movement time. Motion onset is defined as when the hand has moved 2% of the distance to the object. While slightly after actual onset, this threshold accounts for e.g. twitching.

A classifier trained at some point in time also checks if any classifier trained with the same input type at a previous point in time would be more accurate. If so, that one is used instead. This is determined in crossvalidation (section F.3).

F.3 Crossvalidation

Prediction accuracy was calculated for each input at each point of time. This accuracy is defined as the percentage of correctly predicted targets and is calculated separately for 2-object and 3-object trials using within-subject crossvalidation. For each subject, prediction rules are trained using all but one trial from that subject, then tested on the remaining trial. This procedure is repeated as many times as there are trials, with each used as the test trial once.

This form of crossvalidation is different from subject-independent crossvalidation, where classifiers are trained on some subjects and tested on others. Such subject-independent crossvalidation is planned as a future extension of our work.

III. Results

47.5% of 2-object trials reached for object 1 while 52.5% reached for object 2. 30% of 3-object trials reached for object 1, 35% for object 2 and 35% for object 3.

One subject gave the same response to all questionnaire items, so his questionnaire was discarded. One subject's EEG was discarded due to hardware errors. In the other 9 subjects, EEG from 10 (of 360) trials was discarded due to artifacts.

Figs. 3-6 show prediction accuracies for individual modalities (Figs. 3 and 4) and for different modality combinations (Figs. 5 and 6).

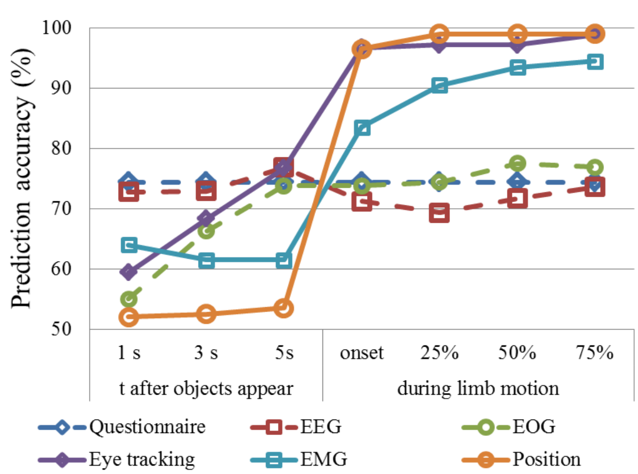

Figure 3. Results for individual sensing modalities in 2-object trials.

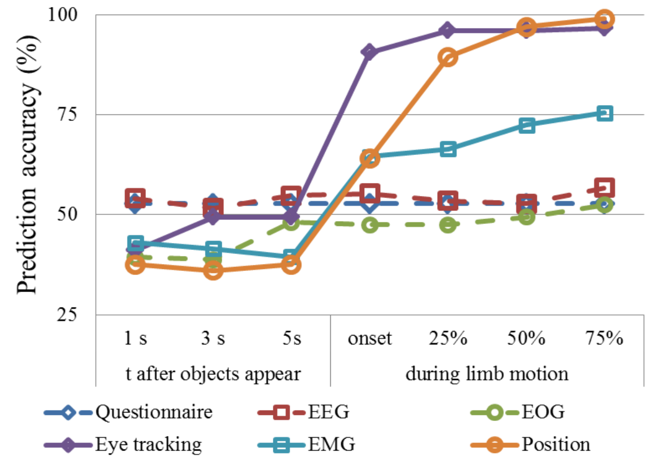

Figure 4. Results for individual sensing modalities in 3-object trials.

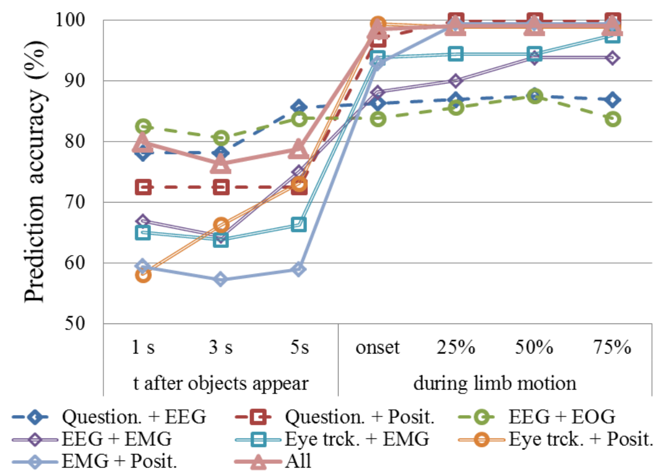

Figure 5. Results for modality combinations in 2-object trials.

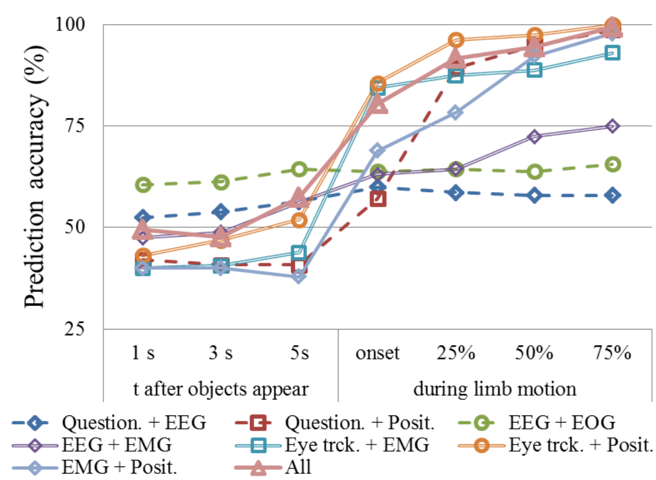

Figure 6. Results for modality combinations in 3-object trials.

IV. Discussion

A. Individual modalities

EEG allows early prediction of the reach target, giving an accuracy of 72.8% for 2-object trials and 54.1% for 3-object trials a second after objects appear. This accuracy is similar to findings in literature: Logar et al. [14] found 75% accuracy when predicting true/false responses while Lee [5] found 40% accuracy when predicting among 4 reach targets.

Eye tracking and *EOG* show gradually increasing accuracy before limb motion onset. At motion onset, eye tracking accuracy increases to over 90% for all trials.

EMG and *hand tracking* exhibit low accuracy before limb motion onset. At motion onset, accuracy increases for both modalities, though accuracy is lower in 3-object trials.

The *questionnaire* has some predictive ability, though it is limited. There are also major intersubject differences, with 2-object accuracies ranging from 50% to 90%.

B. Combining modalities

Combining *questionnaires with EEG* improves accuracy for 2- and 3-object trials, but combining *questionnaires with hand tracking* shows no improvement over a single modality.

Combining *EEG with EOG* improves accuracy over EEG or EOG alone in both 2- and 3-object trials. It thus makes sense to include EOG with EEG in target prediction. Combining *EEG with EMG* shows no improvement over a single modality. This makes sense since EEG yields the highest accuracy before motion onset while EMG is only accurate after onset. Since data from the same time period is used from all modalities, no improvement is possible if individual modalities provide useful data at different times.

Combining *hand tracking with eye tracking* and *EMG with eye tracking* yields lower accuracy than eye tracking alone. However, eye tracking only predicts the target while EMG and hand tracking also measure movement dynamics.

Combining *EMG with hand tracking* shows no benefit over a single modality, likely since both contain similar data.

Combining *all modalities* outperforms any single modality in 2-object trials, but not in 3-object trials. We thus recommend creating classifiers for different times based only on modalities relevant at that time. For example, predictions could first be made using EEG, then switch to hand tracking.

C. Recommendations for use

Some recommendations can be made based on our results and general experience with the hardware.

- For *high accuracy in target prediction*, eye tracking is the optimal choice, especially with many possible targets.

- For *very early prediction*, a combination of EEG and EOG gives the most information prior to limb motion onset.

- If the *subject is unable to perform a movement* (e.g. assistive technologies), eye tracking or EEG and EOG are suitable. EMG may work if some muscle activity remains.

- For *unobtrusiveness*, eye tracking is the only contactless solution in our study. However, hand tracking could be done with contactless systems such as Microsoft’s Kinect.

- For *real-time use*, all examined modalities are suitable, though EEG artifact removal may be difficult to automate.

V. Conclusions

This paper shows the strengths and weaknesses of several sensing modalities. EEG can make early predictions, but is relatively inaccurate. It is suitable for early prediction or for users with motor disabilities. EMG and hand tracking are accurate after motion onset. They are suitable if the user should begin a movement and a device should augment it. Eye tracking is accurate for multiple targets, but gives no information about movement dynamics. It is useful if we wish to impose a motion or bring an object to the user. Finally, a user's preferences can in principle complement any modality, though they may not always be available.

Multimodal data can be difficult to fuse with our approach. An exception was the combination of EEG and EOG, which gives higher accuracy than either individual modality. For better prediction, our algorithm would need to intelligently combine data from different time periods.

References

[1] A. B. Ajiboye, J. D. Simeral, J. P. Donoghue, L. R. Hochberg, and R. F. Kirsch, “Prediction of imagined single-joint movements in a person with high-level tetraplegia,” *IEEE Trans. Biomed. Eng.*, vol. 59, pp. 2755–2765, Oct. 2012.

[2] C. Fleischer and G. Hommel, “A human-exoskeleton interface utilizing electromyography,” *IEEE Trans. Robot.*, vol. 24, pp. 872–882, Aug. 2008.

[3] M. G. Glaholt and E. M. Reingold, “The time course of gaze bias in visual decision tasks,” *Vis. Cogn.*, vol. 17, pp. 1228–1243, Nov. 2009.

[4] I. Krajbich, C. Armel, and A. Rangel, “Visual fixations and the computation and comparison of value in simple choice,” *Nat. Neurosci.*, vol. 13, pp. 1292–1298, Oct. 2010.

[5] E. L. Y. Lee, “Decoding Self-paced Upper Limb Reaching Movements from Brain Signals”, Ph. D. dissertation, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland, 2012.

[6] P. S. Hammon, S. Makeig, H. Poizner, E. Todorov, and V. R. D. Sa, “Predicting reaching targets from human EEG,” *IEEE Signal Process. Mag.*, vol. 69, pp. 69–77, 2008.

[7] S. Waldert, H. Preissl, E. Demandt, C. Braun, N. Birbaumer, A. Aertsen, and C. Mehring, “Hand movement direction decoded from MEG and EEG,” *J. Neurosci.*, vol. 28, pp. 1000–1008, Jan. 2008.

[8] L. Dipietro, A. M. Sabatini, and P. Dario, “Artificial neural network model of the mapping between electromyographic activation and trajectory patterns in free-arm movements,” *Med. Biol. Eng. Comput.*, vol. 41, pp. 124–132, Mar. 2003.

[9] E. A. Corbett, E. J. Perreault, and K. P. Körding, “Decoding with limited neural data: a mixture of time-warped trajectory models for directional reaches,” *J. Neural Eng.*, vol. 9, p. 036002, Apr. 2012.

[10] H. J. Hermens, B. Frerisk, R. Merletti, D. Stegeman, J. Blok, G. Rau, C. Disselhorst-Klug, and G. Hägg, *European Recommendations for Surface Electromyography: Results of the SENIAM project*. Enschede, Netherlands: Roessingh Research and Development, 1999.

[11] P. He, G. Wilson, and C. Russell, “Removal of ocular artifacts from electro-encephalogram by adaptive filtering,” *Med. Biol. Eng. Comput.*, vol. 42, pp. 407–412, May 2004.

[12] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, “A review of classification algorithms for EEG-based brain-computer interfaces,” *J. Neural Eng.*, vol. 4, pp. R1–R13, Jun. 2007.

[13] D. Novak, M. Mihelj, and M. Munih, “A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing,” *Interact. Comput.*, vol. 24, pp. 154–172, May 2012.

[14] V. Logar, A. Belič, B. Koritnik, S. Brezan, J. Zidar, R. Karba, and D. Matko, “Using ANNs to predict a subject’s response based on EEG traces.,” *Neural Netw.*, vol. 21, pp. 881–887, Sep. 2008.