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

Towards a Gaze-Informed Movement Intention Model for Robot-Assisted Upper-Limb Rehabilitation

Vincent Crocher, Ronal Singh, Joshua Newn and Denny Oetomo

Abstract— **Gaze-based intention detection has been explored for robotic-assisted neuro-rehabilitation in recent years. As eye movements often precede hand movements, robotic devices can use gaze information to augment the detection of movement intention in upper-limb rehabilitation. However, due to the likely practical drawbacks of using head-mounted eye trackers and the limited generalisability of the algorithms, gaze-informed approaches have not yet been used in clinical practice.**

This paper introduces a preliminary model for a gaze-informed movement intention that separates the intention spatial component obtained from the gaze from the time component obtained from movement. We leverage the latter to isolate the relevant gaze information happening just before the movement initiation. We evaluated our approach with six healthy individuals using an experimental setup that employed a screen-mounted eye-tracker. The results showed a prediction accuracy of 60% and 73% for an arbitrary target choice and an imposed target choice, respectively.

From these findings, we expect that the model could 1) generalise better to individuals with movement impairment (by not considering movement direction), 2) allow a generalisation to more complex, multi-stage actions including several sub-movements, and 3) facilitate a more *natural* human-robot interactions and empower patients with the agency to decide movement onset. Overall, the paper demonstrates the potential for using gaze-movement model and the use of screen-based eye trackers for robot-assisted upper-limb rehabilitation.

I. INTRODUCTION

Intensive therapy using robotic devices for motor recovery from neurological injuries has been explored for decades [1], [2], [3], [4]. Such robotic devices assist patients with their attempts to perform repeated goal-oriented motor actions. Various forms of interaction methods have also been explored, including patients interacting with on-screen objects [3].

Simultaneously, the absence of spatio-temporal information about the subject's movement (where the person is trying to move to and when they move) drastically limits the potential for assistance and/or correction from a robotic device. The choice of the robotic system is thus often left to either movement-agnostic assistance (*e.g.* deweighting, constant damping/spring) or to impose the timing and movement path (*e.g.* passive movements in position control, see [5] for a complete review) and thus lose the user intention.

To address this issue, researchers have proposed to use gaze tracking to predict motion [1], [2], [6], [7], [8], [9], afforded by the increasing availability and cost of both commercial on-screen and head-mounted eye trackers. Although these works demonstrate the potential for gaze input in the context of rehabilitation, it remains limited as it relies on an explicit or fixed dwell time [2], [10] (fixed time limit is usually given to a user to fixate on a target before movement onset), explicit visual feedback [7] or probability threshold [8] for target selection. In this exact rehabilitation context, Novak et al. [2] showed that the combination of gaze and hand movement information (position) provides the best movement prediction at onset than many other forms of human sensing (*e.g.* EEG, EMG, EOG). Nevertheless, their approach leverages automatic feature extraction and fusion. This has the advantage to allow a fair comparison of multiple techniques but might not generalise well in practice. Indeed, it is impossible to define how much the system will rely on each modality, *i.e.* gaze and/or movement. Thus, the method would likely fail to generalise to people with disrupted movement execution unless an ad-hoc training of the system is performed for each user, which then appears not practical. Novak et al. extended this work, using a screen mounted eye-tracker and decoupling the onset detection (timing) from the spatial detection (target) [11]. Their method, which uses in-game probabilities on top of gaze information, lead to 80% prediction accuracy when evaluated with 12 healthy subjects but requires one full second of data.

Our long-term goal is to develop a *Movement Intention Model* that relies on a hand-eye coordination model by separating the spatial components (obtained from gaze) and the temporal component (obtained from movement initiation) and address the limitations mentioned above. The proposed approach allows the interaction to be dwell-time independent and, therefore, potentially more natural. Further, we believe that both gaze and movement data will be noisy in the context of rehabilitation. Hence, a model that uses gaze input before movement initiation and gaze and movement data after (and also during), will result in robust target selection and refined movement. We also propose treating the gaze information in a probabilistic manner that attempts to reflect subjects' attention to increase its reliability.

This paper presents the first steps in the model development process, in which we evaluate the combination of gaze and movement initiation information to best detect the user intention in reaching actions when assisted by a robot. The explicit separation of spatial and temporal information also allows translating the approach to subjects with movement disorders by not assuming the movement initiation direction. A recent paper [6] presents a similar method, but fusing gaze and EMG after movement initiation while we are interested in the effectiveness of gaze *before* movement onset.

R. S. and J. N. are with the School of Computing and Information Systems, Faculty of Eng. and IT, University of Melbourne, Australia.

V. C. and D. O. are with University of Melbourne and Fourier Intelligence Joint Laboratory, Faculty of Eng. and IT, The University of Melbourne, Australia. (For any correspondence: vcrocher@unimelb.edu.au.)

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

We evaluated our approach with six healthy individuals and achieved a prediction accuracy of 60% and 73% for an arbitrary target choice and an imposed target choice, respectively. From these findings, we expect that the model could 1) generalise better to individuals with movement impairment (by not considering movement direction), 2) allow a generalisation to more complex, multi-stage actions including several sub-movements, and 3) facilitate a more *natural* human-robot interactions and empower patients with the agency to decide movement onset.

II. MOVEMENT INTENTION MODEL

The proposed model aims to provide a target prediction via gaze at the movement onset, allowing the robotic assistant to quickly turn on movement assistance and/or movement correction while leaving the subject's entire action intention. The method assumes that the subject can initiate the movement, but this initiation could be assisted by a movement agnostic support if required, such as a gravity compensation.

We adopt a simple gaze model from existing work on gaze-based intention recognition [12]. This model aims to capture the subject's interest towards a given target (through their visual attention) before movement onset by integrating both saccades towards each target and time spent within them. Following this model, we define fs_i, the *fixation score* for an on-screen target i as a weighted measure between fixation duration and fixation count:

$$fs_i = \lambda \cdot duration_i + (1 - \lambda) \cdot count_i \qquad (1)$$

where $duration_i$ is the total fixation duration on target i, $count_i$ is the number of times a user fixates on i, and $\lambda \in [0, 1]$ is the relative weight given to the fixation duration over fixation count. Because the duration and counts are in different units (time and count, respectively), $duration_i$ will likely be much higher than $count_i$. In practice, we have defined $count_i$ as the number of times i is looked at multiplied by the fixation threshold (200ms), so the two variables are on the same 'scale'. The duration and counts represent two ways of capturing a person's interest in i; the longer a person looks at i and the more number of times the person looks at i, the more the person is interested in i.

The *fixation score* approach to intention recognition requires the user to search for the target of interest actively. In such scenarios, the gaze is actively used by a person to determine the target to reach. The intention to go towards a specific target could be *s*elf-driven or through external stimuli. When self-driven, the user searches the prospective targets, decides by himself/herself the target to reach (forms the intention to reach a specific target), and then initiates the movement. The user could form the intention to reach towards a target with or without searching — they can randomly pick a target even before movement. Alternatively, the intention to move towards a target could be acquired by external stimuli, for example, someone telling the person to search for a specific target. Here, the user has to actively search for the target of interest; therefore, the gaze is likely to be more constrained. For this paper, we aim to understand how the fixation based intention recognition model works under these two scenarios; *self-driven* (free to choose target) and *external stimuli* (constrained to specific targets).

Our work concentrates on target prediction *before* the movement has started (not *after*) and to what extent we can make the on-screen target selection implicit for more natural interaction. That is, we do not rely on an imposed dwell time or visual feedback. We, however, rely on the user movement initiation as sensed by the robotic device. Movement initiation is defined as the $400ms$ before the hand movement reaches a velocity of $0.05m.s^{-1}$. This duration of interest of $d_i = 400ms$ is selected to ensure it encompasses the last fixation present before movement onset [13] and the velocity threshold is set arbitrarily to a low value to ensure that it is met early in the movement and so is achievable even by patients with very little voluntary motion. We denote t_o the movement onset time when the velocity threshold is reached. The fixation score fs_i is thus calculated over the temporal window: $[t_o - d_i, t_o]$. The target with the highest score is selected as the prediction.

III. EXPERIMENTAL VALIDATION

We performed a preliminary experimental validation to investigate the reliability of our proposed approach.

A. Methodology

1) Subjects: We recruited six naive subjects with no history of neurological injury to participate in the experiment. All procedures received approval from The University of Melbourne HEAG (Ethics ID: #1749444.3).

2) Setup: Figure 1 shows our experimental setup. Subjects were seated in front of a touch screen computer with a Tobii 4C eye-tracker attached to the bottom, and their wrist strapped to the EMU robotic device. The EMU is a 3D manipulandum able to generate a force in the three directions to provide assistance, correction or resistance to the subjects' movement. In this experiment, the EMU was used as a position sensor, configured in a transparent mode (*i.e.* to not produce any interaction force), during the reaching movements and to enforce a return to the home reaching position in-between each reaching action. The setup is similar to the setup used in clinical conditions with the EMU device [3] with simply the addition of the eye-tracker device attached to the monitor. Gaze information (gaze point on the screen, measured by the eye-tracker) and movement position (wrist position and velocity measured by the robotic device) were recorded synchronously at 90Hz. Our targets were separated from each other by more than 5cm.

3) Task: After setup, the subjects were presented with three targets on the touch screen in front of them. Subjects were asked to perform 40 reaching movements each. For the first 20 movements, the subjects were presented with three identical targets and left free to reach towards any of them (*Free choice* condition). This setup for the 20 movements was inspired by other works, such as [2], [6]. For the 20 last movements, subjects were constrained to reach one specific

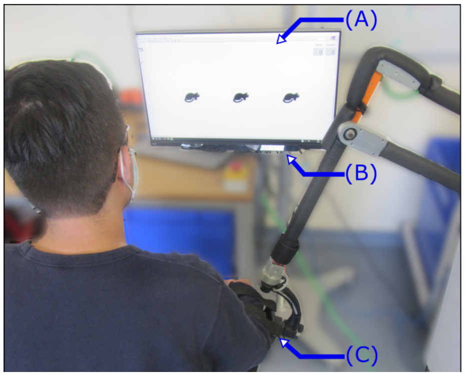

Fig. 1. Experimental setup showing (A) the touch screen with user interface (in *Free choice* mode); (B) the Tobii eye-tracker and (C) the robot cuff with the subject in starting posture.

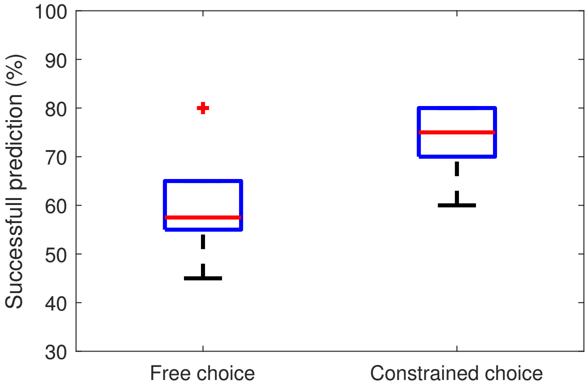

Fig. 2. Successful prediction rate for all subjects and both conditions: *Free choice* (first 20 targets) and *Constrained choice* (last 20 targets).

target among three (one depicting a rabbit among a dog and a cat) (*Constrained choice* condition). Timing of the reaching—initiation and speed—was left free to the subjects.

4) Data analysis: The gaze model (Eq. 1) was applied on gaze position, and hand velocity (obtained through position first-order differentiation) recorded using Matlab (Mathworks). The processing was performed offline but in real-time conditions where the gaze model was used to predict the most likely target before movement onset via a simulation designed using Python programming language.

B. Results

1) Prediction accuracy: Figure 2 presents the distribution of the prediction accuracy in both conditions for all subjects.

On average, the model predicted the correct target in 60% and 73% of the cases, respectively, for the *Free choice* and *Constrained choice* conditions. The results are promising but not as accurate as what some others report. For example [6] report accuracies of around 85% in a situation where the targets were imposed (similar to the *Constrained choice* condition) but also constantly visually presented and using a head-mounted eye-tracker. Not surprisingly, prediction errors were spatially driven, with wrong predictions concentrated around the middle target, as shown in Figure 3.

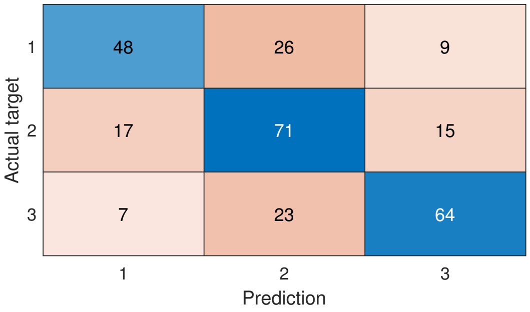

Fig. 3. Confusion matrix of the predictions.

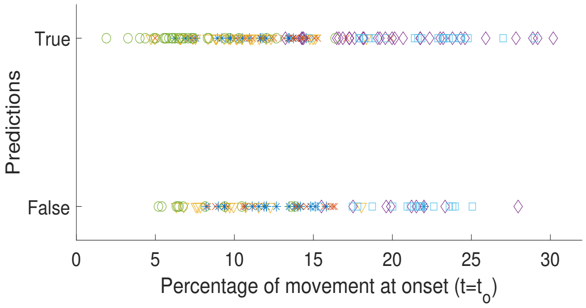

Fig. 4. Distribution of predictions (true *vs* false) against distance at onset ($t = t_o$). Distance is expressed as a percentage of total reaching distance. Each color/symbol depicts a different subject.

2) Active selection impact: Prediction accuracies presented in Figure 2 show that the model better represents the *Constrained choice* condition in which subjects have to find and aim for a given target. This is aligned with the model's intention to capture subjects attention to a target. It is also expected to be more representative of a game-driven rehabilitation scenario than the presentation of three equivalent targets. The gameplay is likely to provide possible movements with different interests or attraction to the user who has to perform an active choice.

In the current experimental setup, it also likely that subjects did not perform any scanning of the target or active selection in the first scenario (*Free choice*) but instead performed a reflex movement towards the screen as soon as the targets appeared.

3) Impact of definition of movement onset: As the model uses a velocity threshold to define the movement onset and that the reaching pace is not imposed on the subjects, the onset time — and hand distance — is variable from one subject to another and even from one reaching cycle to another. Figure 4 shows the impact of the onset distance (as a percentage of reaching total reaching distance) on the model prediction. We see that a prediction happening later on the movement does not favour a higher accuracy, as it could be intuitively suspected. This suggests that a relatively low-velocity threshold and so an early prediction is viable in this setting.

4) Impact of λ: We performed a sensitivity analysis of λ and show the results in Figure 5. Results suggest no

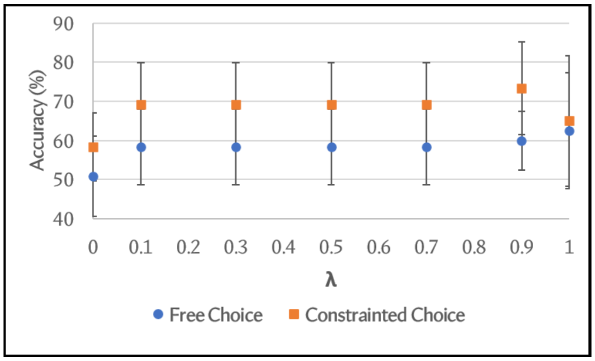

Fig. 5. Sensitivity analysis of λ.

significant difference in prediction accuracy. With different λ values, accuracy (%) for Constrained choices ranged from 58.3 to 73.3, and 50.8 to 62.5 for Free choices.

It is also clearly visible from Figure 5 that the performance is slightly lower for $\lambda = 0$ and $\lambda = 1$. For these values, which correspond to considering only the number of fixations or only the duration of fixations, the prediction scores are closer for both conditions (Constrained and Free). This suggests that an appropriate balance between fixations counts and time helps capturing the actual subjects' attention.

IV. DISCUSSION AND CONCLUSION

This paper investigated a model that combined gaze and movement initiation information to detect the user intention in reaching actions during interaction with a robotic device. Our long-term goal is to develop a *Movement Intention Model* that separates the spatial and temporal components to offer a more natural interaction. The goal is to empower patients with the agency to decide the movement onset.

The prediction accuracy obtained in our experiment in this paper is not yet sufficient for implementation in a neuro-rehabilitation scenario. Nevertheless, the simple model achieved a 2.5 times better prediction than a random choice and showed promise in its possible generalisation with subjects with neurological injuries, compared to a grey-box model including movement direction information used in [2].

Using a head-mounted eye tracker and monitoring fixations to and from the subject's hand could also lead to more satisfying prediction accuracy; the glances at the hand could be used to indicate when the subject is ready to move. It could be easily coupled to a very similar model, but the practical cost remains critical in the application.

Our next steps are to first conduct more studies to understand the full spectrum of gaze behaviours in our context, *e.g.* including saccadic behaviours and pupillary activity [14], and what information these behaviours provide us in determining which target a person has decided to reach. Moreover, the difference in accuracy between *free choice* and *constrained choice* conditions suggests that for gaze-based intention recognition to work, gaze must be explicitly and actively involved during intention formation. For example, more complex game-like tasks may engage participants better in terms of using their gaze.

Furthermore, our proposed model directly relies on the subject's attention to the target to be reached. We can expect that better predictions could be obtained in actual conditions with neurologically impaired subjects due to a higher focus on the task than the current experimental setup.

Finally, it is expected that such a model can be extended to more complex movements and path planning. While the current experiment focused on a prediction before movement onset, the model can be applied over a continuous-time (in a windowed fashioned) to predict directions and re-planning changes. This would allow for more realistic scenarios, including multi-stage movement and functional tasks practice, which are currently lacking in robot-assisted rehabilitation.

REFERENCES

[1] A. Frisoli, C. Loconsole, D. Leonardis, F. Banno, M. Barsotti, C. Chisari, and M. Bergamasco, "A new gaze-bci-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 1169–1179, 2012.

[2] D. Novak and R. Riener, "Enhancing patient freedom in rehabilitation robotics using gaze-based intention detection," in *13th IEEE International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 2013.

[3] J. Fong, V. Crocher, Y. Tan, D. Oetomo, and I. Mareels, "Emu: A transparent 3d robotic manipulandum for upper-limb rehabilitation," in *2017 International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 2017, pp. 771–776.

[4] J. Ibáñez, E. Monge-Pereira, F. Molina-Rueda, J. I. Serrano, M. D. Del Castillo, A. Cuesta-Gómez, M. Carratalá-Tejada, R. Cano-de-la Cuerda, I. M. Alguacil-Diego, J. C. Miangolarra-Page *et al.*, "Low latency estimation of motor intentions to assist reaching movements along multiple sessions in chronic stroke patients: a feasibility study," *Frontiers in neuroscience*, vol. 11, p. 126, 2017.

[5] A. Basteris, S. M. Nijenhuis, A. H. Stienen, J. H. Buurke, G. B. Prange, and F. Amirabdollahian, "Training modalities in robot-mediated upper limb rehabilitation in stroke: a framework for classification based on a systematic review," *Journal of NeuroEngineering and Rehabilitation*, vol. 11, no. 1, p. 111, Jul. 2014.

[6] N. Krausz, D. Lamotte, I. Batzianoulis, L. Hargrove, S. Micera, and A. Billard, "Intent prediction based on biomechanical coordination of emg and vision-filtered gaze for end-point control of an arm prosthesis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2020.

[7] B. J. Hou, P. Bekgaard, S. MacKenzie, J. P. P. Hansen, and S. Puthusserypady, "Gimis: Gaze input with motor imagery selection," in *ACM Symposium on Eye Tracking Research and Applications*, 2020.

[8] H. Zeng, Y. Shen, X. Hu, A. Song, B. Xu, H. Li, Y. Wang, and P. Wen, "Semi-autonomous robotic arm reaching with hybrid gaze–brain machine interface," *Frontiers in Neurorobotics*, vol. 13, p. 111, 2020.

[9] E. M. Young, T. J. Withrow, and N. Sarkar, "Design of intention-based assistive robot for upper limb," *Advanced Robotics*, vol. 31, no. 11, pp. 580–594, 2017.

[10] J. A. Díez, J. M. Catalán, L. D. Lledó, F. J. Badesa, and N. Garcia-Aracil, "Multimodal robotic system for upper-limb rehabilitation in physical environment," vol. 8, no. 9.

[11] R. Riener and D. Novak, "Movement onset detection and target estimation for robot-aided arm training," vol. 63, no. 4, pp. 286–298, publisher: De Gruyter Oldenbourg Section: Automatisierungstechnik.

[12] R. Singh, T. Miller, J. Newn, E. Velloso, F. Vetere, and L. Sonenberg, "Combining gaze and ai planning for online human intention recognition," *Artificial Intelligence*, p. 103275, 2020.

[13] J. L. Vercher, G. Magenes, C. Prablanc, and G. M. Gauthier, "Eye-head-hand coordination in pointing at visual targets: spatial and temporal analysis," *Experimental Brain Research*, vol. 99, no. 3, pp. 507–523, Jan. 1994.

[14] Y.-M. Jang, R. Mallipeddi, and M. Lee, "Identification of human implicit visual search intention based on eye movement and pupillary analysis," *User Modeling and User-Adapted Interaction*, vol. 24, no. 4, p. 315344, Oct. 2014.