Towards Integrating Eye Gaze Tracking into a Multimodal Dialog Agent for Remote Patient Assessment

Daniel Tisdale, Jackson Liscombe, David Pautler, and Vikram Ramanarayanan

Introduction

We demonstrate a prototype that integrates automated eye gaze tracking into an already existing multimodal conversational platform for remote patient assessment and monitoring (**NEMSI**; short for "**NE**urological and **M**ental health **S**creening **I**nstrument"). The platform engages patients in an interactive dialog session and guides patients through several spoken, orofacial, cognitive, and gaze tasks inspired by clinical protocols. Novel additions to the dialog protocol include a selection of exercises that have been widely used in oculomotor pathology research as well as clinical practice, including: smooth pursuit, saccade, free image exploration, directed image exploration, and the congruent and incongruent Stroop tests.

System Design

NEMSI

- A multimodal conversational platform for remote patient diagnosis and monitoring, which extracts a variety of biomarkers after engaging patients in an interactive dialog session
- The conversational assessment protocol can include a customizable subset of following tasks, depending on the nature of the disease in question.

Eye Gaze Tracking

- We use Webgazer.js as our eye gaze tracker in NEMSI. WebGazer.js has two key components: a pupil detector and a gaze estimator using regression analysis informed by user interactions. Webgazer.js was originally designed for use in evaluating user interaction with websites and, as such, by default uses feedback from user mouse movements for continuous gaze calibration. Because many of our users suffer from diseases that may affect motor control and impair their use of a mouse, we have turned off this Webgazer.js feature and instead start with just one mouse-related calibration task, described below.
- The NEMSI virtual agent can engage the user in the following tasks:

Figure 1. A **Calibration** task in which the user is asked to click circles on the border of their screen as the look at them. The results are used to inform the eye gaze tracker to better estimate user gaze position for the remainder of the session.

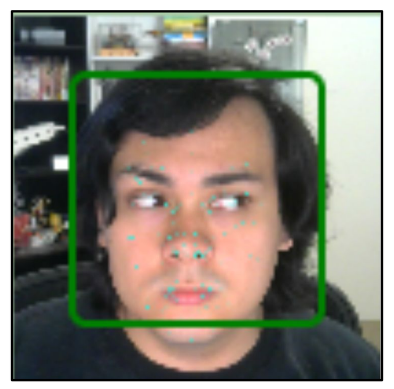

Figure 2. **Extreme Vertical and Horizontal Eye Gaze Tasks** in which the user is asked to gaze as far to the left, right, up, and down as they can, both slowly and rapidly.

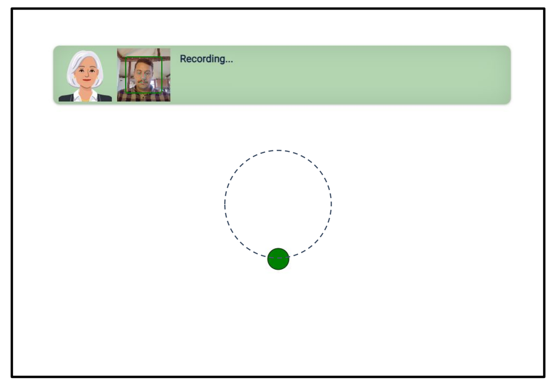

Figure 3. **Smooth Pursuit Tasks** in which the user is asked to follow a moving circle with their eyes while it moves in either a line or a circle.

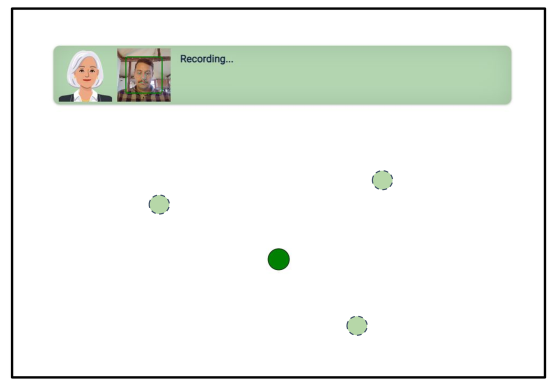

Figure 4. A **Saccade Task** in which users are asks to direct their gaze to dots that appear briefly on the screen in random locations.

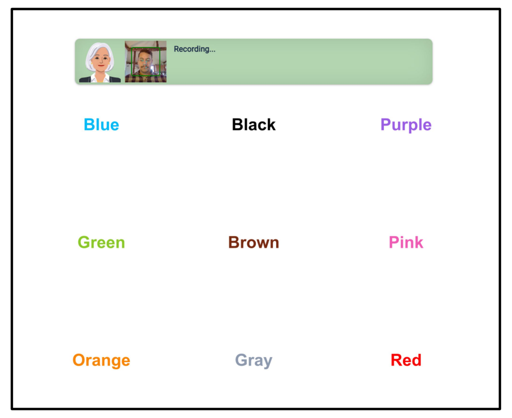

Figure 5. **Modified Congruent Stroop Color and Word Tests (SCWT)** in which the user is presented with a matrix of color words whose text color matches their semantic meanings and is asked to either read the words in order or to find a specific word.

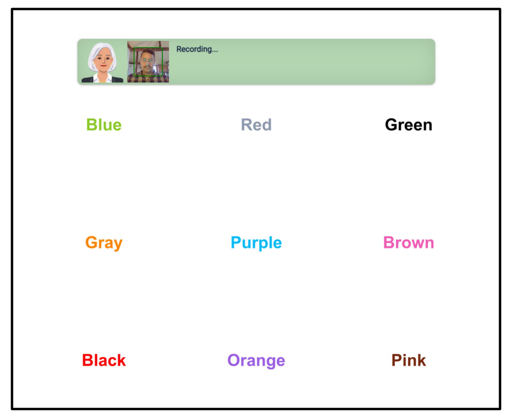

Figure 6. **Modified Incongruent SCWTs** identical to the ones above, but in which the text color of the words do not match their semantic meanings (e.g. the word "brown" with an ink color of pink).

Figure 7. **Directed Image Exploration** in which the user is tasked with finding different objects within the image

Figure 8. **Undirected Image Exploration** in which the user is prompted to freely look at and explore the image

Verification & Results

To verify that our observed error rate matched that reported by WebGazer.js developers, we designed a "shrinking dot task" in which internal testers were asked to use their eye gaze to follow a smoothly moving and shrinking dot as it moved and bounced off the edges of their computer screens. We found the mean accuracy for the task fell below 50% once the dot was smaller than 180 pixels (px), on average. This result falls within the mean error reported of 210.6px (SD=86.3px). This error of 180px has informed our task user interface design that set fixation targets at least 200px apart from one another.

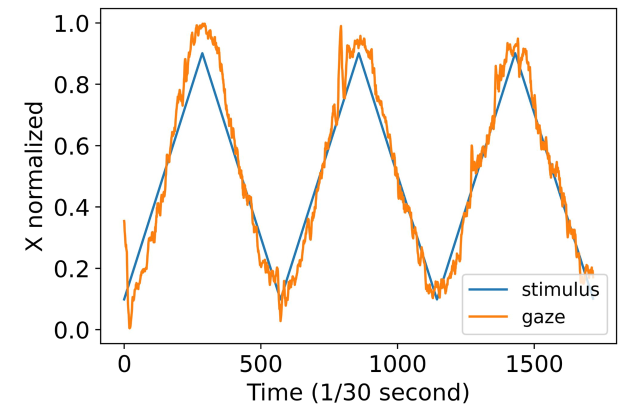

Figure 9. Smooth Pursuit Task Example (X coordinates)

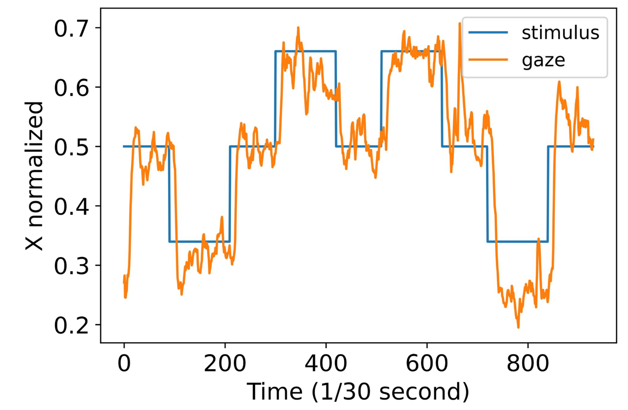

Figure 10. Saccade Task Example (X coordinates)

Task	Error X (%)		Error Y (%)	
	Mean	SD	Mean	SD
Smooth Pursuit Line	8.8	3.7	23.4	6.6
Smooth Pursuit Line	11.2	4.9	21.2	8.5
Saccade	7.5	5.3	19.6	4.1
Congruent SWCT	13.7	10.8	23.6	10.6
Incongruent SWCT	14.2	12.6	21.1	8.3
Directed Image	14.8	12.8	20.7	11.8

Table 1. Eye Gaze Prediction Errors

We also analyzed eye tracking data of 13 healthy controls who completed all aforementioned tasks. Users had screen sizes of varying dimensions and had all their eye gaze coordinates normalized to the range [0,1] based on their respective screen dimensions. Figures 9 and 10 show two examples of the x coordinate eye gaze data obtained from a single user in the smooth pursuit and saccade tasks, respectively. In addition to visual inspection, we also computed fixation errors over all tasks and subjects. Table 1 lists average x and y displacement error per task, as percent of screen size. Given average screen size (858 x 1610px), and considering 180 pixel average error, expected displacement error for healthy controls should be ≤11.2% screen width and ≤21.0% screen height. As a reminder, our system is not using mouse information for these tasks, which limits the eye gaze prediction accuracy when compared to Webgazer. Along with this, some tasks require a still unknown amount of reaction time for healthly users' gaze to reach different targets on the screen that have not been accounted for when calculating these aggregate error rates. All in all, these data and analyses on healthy controls provides us with a successful proof-of-concept towards our next step: investigating the feasibility and utility of deploying this technology to analyze data from patients with cognitive and neurological disorders.

Conclusions

- We have demonstrated how to incorporate the eye gaze modality into NEMSI, a multimodal conversational platform for remote patient diagnosis and monitoring, which extracts speech, facial, cognitive, respiratory and now eye-gaze-based biomarkers while engaging patients in an interactive dialog session.

References

- A. Papoutsaki, P. Sangkloy, J. Laskey, N. Daskalova, J. Huang, and J. Hays, "WebGazer: Scalable webcam eye tracking using user interactions," in Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI-16). AAAI, 2016, pp. 3839–3845.
- F. Scarpina and S. Tagini, "The stroop color and word test," Frontiers in Psychology, vol. 8, 2017. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00557