Feasibility of Head-Mounted OCR-Based Assistive Technology for Dyslexic Students

Muhammad Haj Ali   Ali Shaer

February 25, 2025

1 Introduction

Dyslexia affects approximately 5% to 15% of the global population [1], presenting significant challenges in academic settings where students struggle with accurate word recognition, poor spelling, and slow reading [2]. These difficulties can hinder academic performance but do not reflect intelligence or a lack of motivation to learn [3] [4]. Traditional assistive technologies, particularly Optical Character Recognition (OCR) and Text-to-Speech (TTS), have shown promise in addressing these challenges. For instance, applications like Augmenta11y employ real-time OCR with customizable settings for font size and spacing [5]. However, current solutions are typically limited to mobile devices, which present practical challenges in classroom settings, as students must constantly hold and position their phones to capture text. In contrast, a wearable solution provides hands-free operation and automatically aligns with the user's field of view, allowing students to naturally follow the teacher's writing on the whiteboard.

In this research, we present a new approach to improve accessibility for individuals with dyslexia and test its efficacy. We developed an application built upon eye-tracking glasses that captures images from the user's field of view, detects whiteboards in the captured images, performs OCR on the detected whiteboard content, and displays the output both in audio and textual format in real-time. Additionally, our application captures the user's eye gaze behavior, providing a testbed for uncovering reading behavior patterns related to dyslexia through eye gaze analysis. Our study focuses on analyzing the OCR system's accuracy and response times in classroom settings. We present the system architecture, performance metrics, and considerations for future user studies.

2 System Overview

We developed an application to detect whiteboards and extract text from them, allowing for auditory playback of the detected text and a notebook to view and edit the content. The application was built using Pupil Labs Neon eye-tracking glasses, a wearable platform that combines real-time scene capture and eye tracking in a lightweight design resembling regular eye-wear. A mobile device running the Pupil Labs companion application processes and streams the scene capture and eye-tracking data via a WebSocket server to our Python application. The Python application interfaces with the GPT-4V model using the prompt: "You are in a classroom with a whiteboard, detect the whiteboard, perform OCR on the text written on it", managing whiteboard detection, OCR, and TTS-1 model for text-to-speech. The main processing occurs on a Windows 11 laptop with an Intel i7-12700H processor (2.3GHz), 16GB RAM, and an NVIDIA GeForce RTX 3050 Ti GPU. A dedicated capture button starts the process. During runtime, GPT processes images captured by Neon's camera, returning extracted text along with a text-to-speech .wav file (see Figure 1).

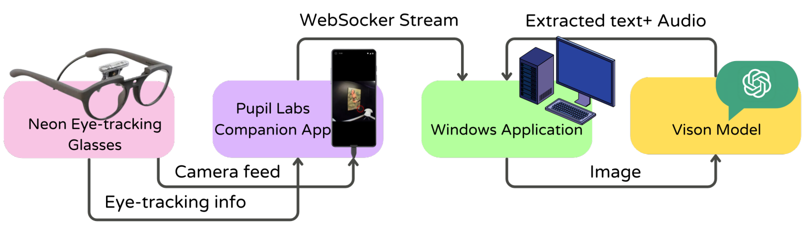

Figure 1: System Architecture and Data Flow Pipeline

Different user scenarios are likely to require different solutions. Thus, our application was also designed to showcase content in audio format and customizable text display. Based on user study findings from Augmenta11y [5], where dyslexic readers expressed preferences for customizable text presentations, our system implements these accessibility features, allowing users to modify text size, font type, text color, and background color. (see Figure 2) to accommodate individual reading preferences, which can be tested through a user study in future research. For research and application evaluation purposes, we implemented a flexible monitoring interface that displays the camera feed and debug console (see Figure 2), this interface allows researchers to observe system behavior in real-time and identify potential issues. All interface components - including the camera feed, console output, and text display - can be resized or hidden based on research requirements or user needs.

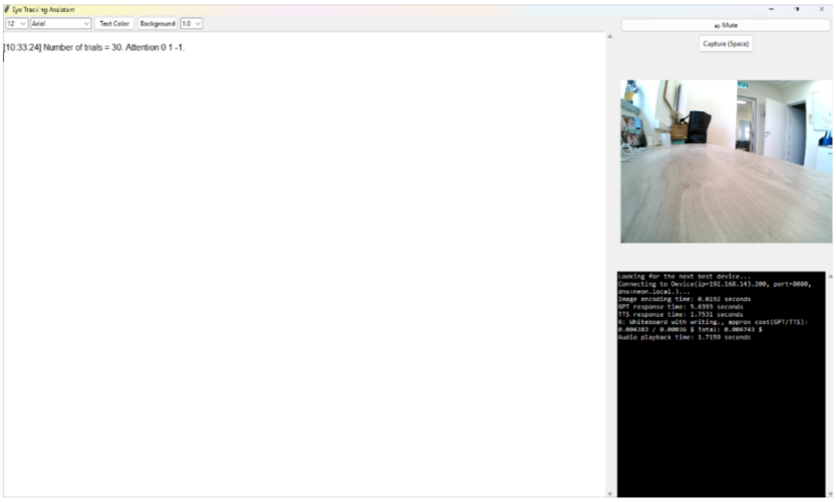

Figure 2: System architecture showcasing data flow from the Neon glasses to the processing unit.

Our application was designed with future user studies in mind. The application tracks key performance parameters related to the OCR pipeline, including image transfer latency, whiteboard detection time, and text extraction processing duration. For eye gaze analysis, the Neon eye-tracking glasses capture comprehensive data at high temporal resolution: gaze points (200 Hz), head motion data (110 Hz), and pupillometry data including eye position and pupil diameter (200 Hz).

3 Experiment

We conducted a controlled experiment to evaluate our system's performance across multiple dimensions: accuracy, latency, and usability. The experiment took place in a standard classroom setting with controlled lighting conditions and predefined test patterns. The test configuration involved viewing a 1m $\times$ 1m whiteboard from 2-meters and three viewing angles ($-30°, 0°, +30°$ measured from perpendicular to the board). For test material, we used common English words written in standard handwriting with black dry-erase markers (see Figure 3), conducting 5 tests per condition. The classroom environment was maintained at consistent lighting levels of 160-175 lux, measured using an MT-912 Light Meter, with care taken to avoid direct sunlight or shadow interference.

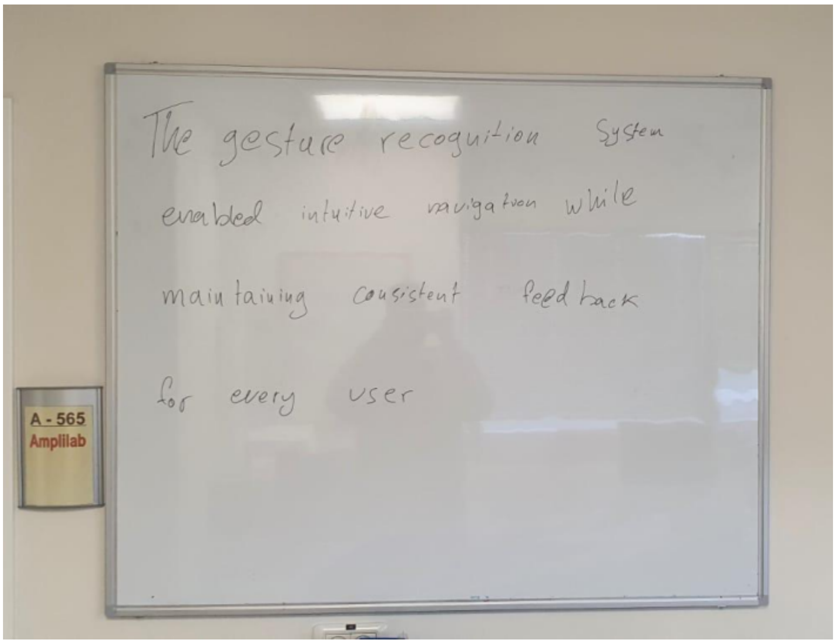

Figure 3: Sample Test Environment- Whiteboard Text Capture at 0° Viewing Angle

To evaluate system performance, we focused on two key metric categories: recognition accuracy and system latency. Recognition accuracy was measured through word recognition rate, calculating the percentage of correctly identified words, while system latency was broken down into image encoding time, GPT response generation time, text-to-speech conversion time. Each component was measured in seconds and logged for analysis. Cost metrics were also tracked to assess the system's operational expenses. Each combination of distance and angle was tested across all trials, resulting in 15 total test cases (1 distance × 3 angles × 5 repetitions).

4 Preliminary Results

The image encoding time remained constant at 0.0045 seconds across all viewing angles, indicating robust streaming connection from neon, GPT response generation showed consistent performance, with averages ranging from 3.31 to 3.44 seconds across angles. Notably, at 0° (perpendicular view), GPT response time averaged 3.31 seconds (range: 2.73-3.77s), while at -30° and

+30°, it averaged 3.44 and 3.33 seconds respectively (See Table 1), but occasional delays were observed at +30°, due to the system pausing to process additional spaces and punctuation marks detected in the image. These processing times, often exceeding 3 seconds, highlight a critical need for optimization in future iterations to achieve real-time performance necessary for effective seamless classroom use. OCR accuracy remained remarkably high across all viewing angles (see Figure 4), with the system achieving 97.3% accuracy at 0°, 96.0% at -30°, and 100% at +30°. These results suggest that while processing times may vary, the system maintains high recognition accuracy regardless of viewing angle, with standard deviations of ±6.0%, ±3.7%, and ±0.0% respectively.

Table 1: Average Response Time Breakdown Across Different Viewing Angles

Viewing Angle	Image Encoding (s)	GPT Response (s)	TTS Conversion (s)
0°	0.0045	3.31 (2.73-3.77)	1.91 (1.72-2.04)
−30°	0.0045	3.44 (3.02-3.77)	2.10 (1.90-2.38)
+30°	0.0045	3.33 (2.96-3.96)	3.70 (1.29-4.60)

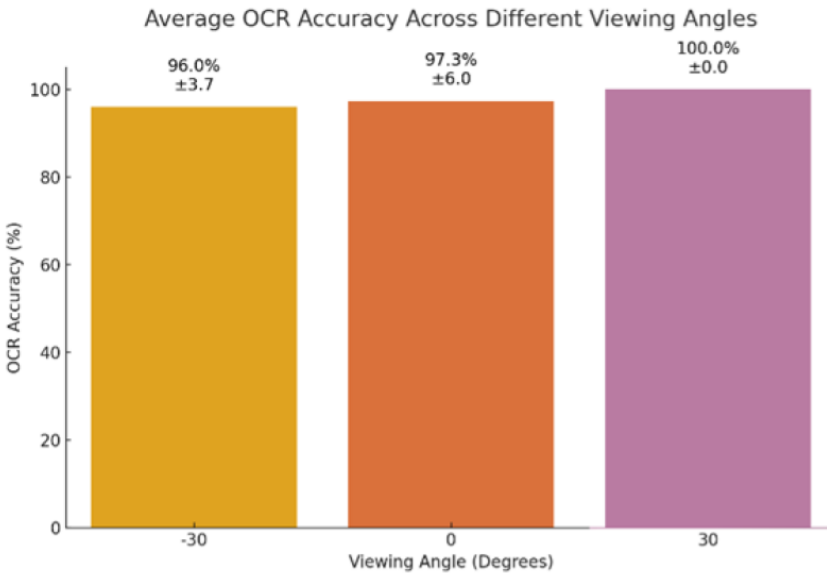

Figure 4: Average Response Time Breakdown Across Different Viewing Angles (-30°, 0°, +30°)

5 Future Work and Discussion

Our experimental results demonstrated high OCR accuracy across different viewing angles. However, several technical limitations remain:

- The current implementation is limited to digital text recognition and does not support handwritten content or diagrams.
- The system's cost, including Pupil Labs Neon eye-tracking glasses and ongoing GPT-4V API expenses, hinders widespread educational adoption.
- Processing delays occasionally exceed 3 seconds, highlighting a need for optimization.

Future improvements should include implementing adaptive contrast adjustments and noise reduction techniques to enhance text recognition. Additionally, exploring offline OCR models such as Tesseract could reduce dependency on API costs, albeit at a potential trade-off in accuracy. Developing specialized models for handwritten content would further improve real-world classroom applicability.

6 Conclusion

We developed an assistive technology system combining eye-tracking, OCR, and text-to-speech capabilities to support dyslexic students in classroom environments. Our experimental results demonstrated high OCR accuracy (96-100%) across different viewing angles but highlighted latency challenges that require further optimization. While cost and handwritten content recognition remain limitations, our system lays a foundation for future advancements in educational accessibility technology.

References

[1] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. 2013. https://doi.org/10.1176/appi.books.9780890425596

[2] Athanaselis, T., et al. "Making assistive reading tools user friendly: A new platform for Greek dyslexic students empowered by automatic speech recognition." *Multimedia Tools and Applications*, 68(3), 2014. https://doi.org/10.1007/s11042-012-1073-5

[3] Tunmer, W. E., & Greaney, K.T. (2010). Defining dyslexia.*Journal of Learning Disabilities*, 43(3), 229–243.

[4] Rello, L., Baeza-Yates, R. "Dyslexia and web accessibility." *12th Web for All Conference*. 2015. https://doi.org/10.1145/2745555.2746655

[5] Gupta, T., Aflatoony, L., & Leonard, L. (2021). *Augmenta11y: A Reading Assistant Application for Children with Dyslexia*. In The 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '21). https://doi.org/10.1145/3441852.3476530

[6] De Luca, M., Borrelli, M., Judica, A., Spinelli, D., & Zoccolotti, P. (2002). *Reading words and pseudowords: An eye movement study of developmental yslexia.*Brain and Language, 80(3), 617–626. https://doi.org/10.1006/brln.2001.2637

[7] Rauschenberger, M., Baeza-Yates, R., & Rello, L.(2019).Technologies for Dyslexia. in *Springer Handbook of Accessibility.* https://doi.org/10.1007/978-1-4471-7440-0_31