Cover Page

Title: Eye Tracking during Passage Reading Supports Precise Oculomotor Assessment in Ataxias

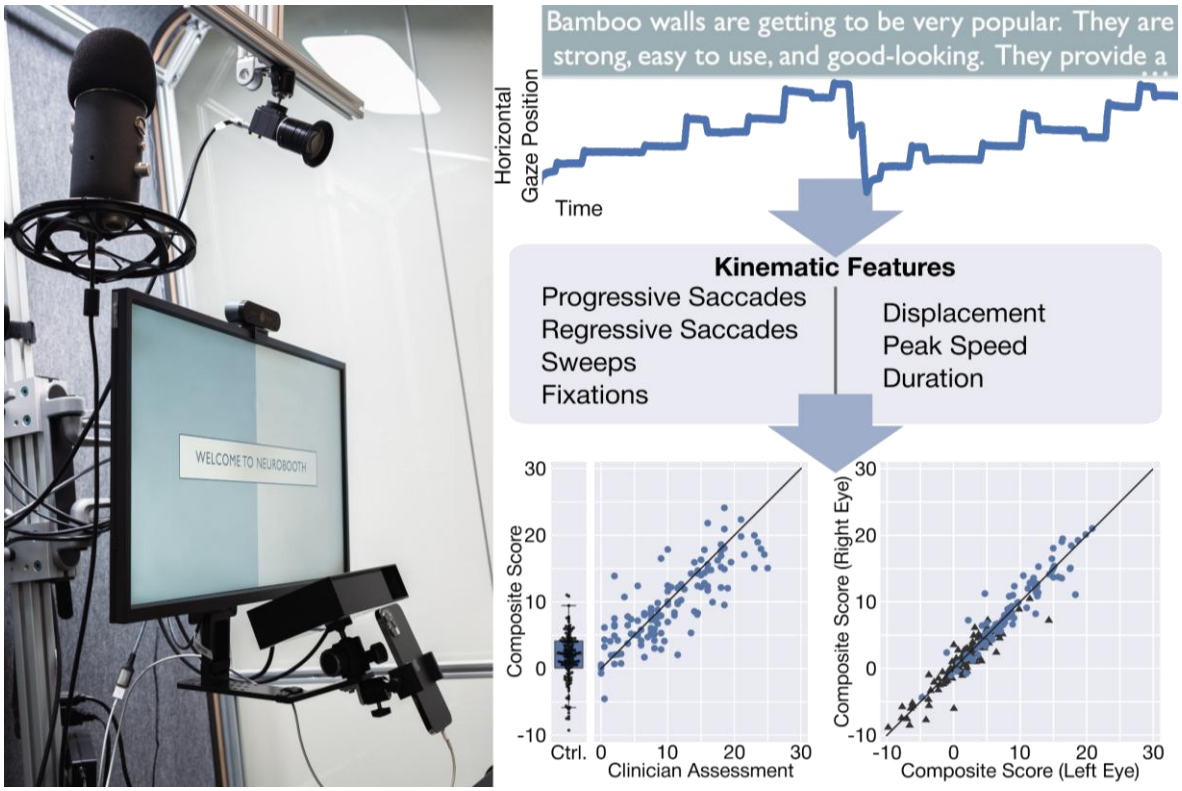
Running Title: Oculomotor assessment during reading

Authors: Brandon Oubre[1,2], Faye Yang[2], Anna C. Luddy[2], Rohin Manohar[2], Nancy N. Soja[2], Christopher D. Stephen[1,2,3], Jeremy D. Schmahmann[1,2,3], Divya Kulkarni[1,2], Lawrence White[2], Siddharth Patel[1,2], Anoopum S. Gupta[1,2,3]

Affiliations:
[1] Harvard Medical School
[2] Massachusetts General Hospital Department of Neurology
[3] Massachusetts General Hospital Department of Neurology, Ataxia Center

Corresponding Author: Anoopum S. Gupta

Thumbnail



Abstract

Abnormal eye movements occur early in the course of disease in many ataxias. However, clinical assessments of oculomotor function lack precision, limiting sensitivity for measuring progression and the ability to detect subtle early signs. Quantitative assessment of eye movements during everyday behaviors such as reading has potential to overcome these limitations and produce functionally relevant measures.

In this study, we analyze eye movements in individuals with ataxia during passage reading. Binocular gaze sampled at 1000 Hz was collected from 102 individuals with ataxia diagnoses (including 36 spinocerebellar ataxias, 12 Friedreich's ataxia, and 5 multiple system atrophy among other conditions) and 70 healthy controls participating in the Neurobooth study. Longitudinal data were available for 26 participants with ataxia. Saccades were categorized as progressive (rightward) saccades, regressive saccades, or sweeps (large displacement saccades primarily generated when scanning to the beginning of the next line) based on their direction and displacement. Saccade and fixation kinematics were summarized using 28 statistical features. A linear model was trained to estimate clinician-performed ataxia rating scale scores.

Model scores were reliable (ICC=0.96, $p<0.001$) and demonstrated convergent validity with Brief Ataxia Rating Scale total ($r=0.82$, $p<0.001$), oculomotor ($r=0.52$, $p<0.001$), and speech ($r=0.73$, $p<0.001$) scores, as well as patient surveys. The scores were also sensitive to disease progression ($d=0.36$, $p=0.03$), demonstrated strong separability between healthy controls and participants with ataxias (AUC=0.89, $p<0.001$), and showed evidence of the ability to detect subclinical oculomotor patterns (AUC=0.69, $p=0.02$). Several kinematic saccade and fixation features demonstrated strong differences across disease severity groups. Notable features included the mean angular displacement of fixations ($\eta^2=0.44$, $p<0.001$), the number ($\eta^2=0.27$, $p<0.001$) and frequency of saccades ($\eta^2=0.25$, $p<0.001$), and the proportion of regressive saccades ($\eta^2=0.11$, $p<0.001$).

Quantitative assessment of eye movements during passage reading were highly informative of ataxia severity, were sensitive to disease progression, and enabled detection of subclinical signs. These properties support the inclusion of video-oculography-based measures of reading in natural history studies and clinical trials. Furthermore, this study demonstrates the feasibility of integration of oculomotor assessments in clinical workflows.

Introduction

Cerebellar ataxia is a neurologic phenotype arising from impaired function of the cerebellum and connected structures.[1,2] Ataxia can result from diverse etiologies, including hereditary and sporadic disorders that are neurodegenerative and cause progressive disability over time.[1] Symptoms include balance and gait difficulties, clumsiness of movements, and impaired speech (dysarthria),[1] resulting in limited autonomy and reduced quality of life.[3,4] Furthermore, ataxias are frequently accompanied by eye movement abnormalities[5,6] due to the involvement of the cerebellum in oculomotor control.[7] These abnormalities manifest early in the course of disease progression[5] and are diagnostically informative.[6]

Accelerating drug development efforts for ataxias underscore the need for sensitive and meaningful measures of disease progression, particularly in the early stages of disease where disease-modifying treatments may provide the most benefit.[8–10] However, standard practice currently relies on semiquantitative clinical rating scales, such as the Brief Ataxia Rating Scale[11] (BARS), Scale for Assessment and Rating of Ataxia[12] (SARA), and International Cooperative Ataxia Rating Scale[13] (ICARS). These scales either do not or only coarsely assess oculomotor function.[14] Furthermore, because of the speed and small spatial displacement of eye movements, reliance on human visual perception fundamentally limits the ability of clinician-rated scales to precisely assess the subtle characteristics of abnormal eye movements and their evolution over time.

Digital technologies have great potential to address limitations of traditional clinical rating scales across different motor domains.[15–18] In particular, video oculography is a powerful tool for understanding oculomotor and cognitive impairment across a variety of neurologic diseases.[14,19] Though there is value in the application of video oculography to pure oculomotor tasks designed to elicit pathologic signs,[14] the analysis of eye movements during a natural task such as reading offers advantages. Reading typically generates more saccades—and more varied saccades—than a prosaccade task of similar duration. Reading is also a common everyday behavior with functional relevance[20–22] and may therefore be suitable for passive at-home assessments in the future using smart glasses or front-facing cameras on mobile devices. The increased complexity and coordination required by reading (in contrast to pure oculomotor tasks) may also help reveal additional clinically informative patterns of eye movements.[23] Though reading aloud is a less common everyday behavior in comparison to silent reading and elicits differences in eye movement patterns (including increased fixation duration and smaller saccade displacements),[24] it presents additional opportunities for the concurrent analysis of coordination between abnormal eye movements and dysarthric speech.[25]

In this work, we hypothesized that analysis of eye movement kinematics during passage reading would enable precise and reliable assessment of oculomotor abnormalities in a diverse ataxia population. We show that the reading-based assessment reliably quantifies disease severity in a more continuous manner than ataxia rating scales, and also demonstrates sensitivity to disease progression and detection of subclinical oculomotor signs. We validate this digital assessment in a relatively large, heterogeneous ataxia cohort with age-matched healthy controls, and

demonstrate the practicality of routinely administering such an assessment in clinical settings. These findings highlight the rich information contained in the patterns of eye movements during reading and demonstrate the potential benefit of integrating video oculography-based assessments into clinical practice and clinical research.

Materials and methods

Data Collection

The data used in this analysis were obtained from the Neurobooth[26] study. The Neurobooth is situated in the neurology clinic at Massachusetts General Hospital and supports time-synchronized data collection using a large array of sensors to capture performance on a variety of behavioral stimuli. The study aims to collect longitudinal digital phenotypic data in a relatively large and diverse cohort of individuals with neurologic disorders. Participants in the study typically visit the booth immediately prior to or after their scheduled appointments in the clinic, at which time the neurologist-administered outcome assessments are documented. The task battery for the Neurobooth study, which included the passage reading task, is designed to be completed within 30-40 minutes.

Participants were seated inside the booth (Framery Model Q) on a full-back chair with a standard headrest (Fig. 1A). Some participants with limited mobility opted to use their wheelchair instead of the provided chair. Stimuli were presented on a 24-inch BenQ monitor with a resolution of 1920 by 1080 pixels and a refresh rate of 240 Hz. Participants were seated roughly 58 cm (57.8 ± 3.9 cm; mean ± standard deviation) from the monitor. The monitor height was adjustable, though to maintain throughput on busy days it was only adjusted (such that the top of the monitor was at eye level) for particularly tall or short participants. An EyeLink Portable Duo (SR Research Ltd.) was mounted on a bracket attached to the monitor mount such that the center of the eye tracker was 12 cm in front of and vertically level with the bottom of the screen (Fig. 1A). A target sticker with concentric black and white circles was affixed to the center of participants' foreheads during data collection. Binocular gaze data were collected at 1000 Hz using the Portable Duo's head-free configuration (i.e., without the use of a chinrest).

Eye tracking calibration was performed using a five-point stimulus at the beginning of the Neurobooth task battery. In some situations, the calibration was not able to be validated due to severe disease signs (e.g., saccadic intrusions). After calibration, participants completed a battery of oculomotor (e.g., saccades, smooth pursuit, fixation) and speech (e.g., sustained phonation, repeated consonants) tasks before being presented with the passage reading task. Participants were presented with a brief instruction video, and then shown the Bamboo passage[27] (Fig. 1B). Participants took as much time as needed to read the passage aloud. A research coordinator concluded the recording when the participant was finished reading.

Study Participants

Analyses were performed on data collected between April 28, 2022 and May 2, 2024. This data set included 139 individuals with clinically and/or genetically diagnosed ataxias and 78 healthy neurologically age-matched healthy controls. Controls were classified as MGH controls or general population controls based on the recruitment mechanism. MGH controls were the family and caregivers of patients who received care at Massachusetts General Hospital Department of Neurology. General population controls were recruited more broadly using flyers, the Neurobooth website, and advertisements on the Rally for Partners platform. The distinction between controls was made because recruitment from the general population may result in inclusion of some individuals who participate due to a concern about their neurological function.[28–30]

A total of 45 participants (20.7%, 8 controls, 37 with ataxia) were subsequently excluded from analysis because of poor eye tracking data quality using the analytic criteria described in the Data Processing Methods and Supplemental Material. Statistical comparisons were used to compare the clinical characteristics of included versus excluded participants (see Supplementary Results). The demographics of the remaining 70 healthy controls and 102 participants with ataxia included in analyses are detailed in Table 1. There was no statistical difference (AUC=0.50, p=0.95) in the ages of the remaining healthy controls (54.9 ± 18.3 years old) and participants with ataxia (55.2 ± 16.5 years old). Longitudinal data were present for 37 controls and 26 individuals with ataxia. Consecutive visits were 85–623 (235.9 ± 93.2) days apart. The six most frequent diagnoses were Spinocerebellar Ataxia type 3 (SCA-3, N=19); Friedreich's Ataxia (FRDA, N=12); SCA-6 (N=9); Cerebellar Ataxia, Neuropathy and Vestibular Areflexia Syndrome (CANVAS, N=7); SCA-2 (N=6); and the cerebellar subtype of Multiple System Atrophy (MSA-C, N=5).

All participants provided informed consent and the research protocol was approved by the Massachusetts General Hospital Institutional Review Board (2021P000257, approved March 3, 2021).

Contextual Clinical Information

Participants with ataxia were assessed using the Brief Ataxia Rating Scale[11] (BARS) during their clinical appointments, which has a total score range of 0–30 with 30 being the most severe. BARS subscores (either finger-nose or heel-shin) were missing for four visits corresponding to three participants. One participant was missing subscores worth eight points and the remaining participants were missing subscores worth four points. BARS total scores were normalized to account for missing subscore values. For the 102 participants with ataxia, normalized BARS ranged from 0–25 (10.6 ± 6.5) points. Fourteen participants had a BARS oculomotor score of 0 at the time of assessment, indicating no clinically-observed oculomotor signs during that assessment. (Three of these participants had increased BARS oculomotor scores on subsequent assessments.)

All participants were asked to complete a battery of patient-reported outcome measures (PROMs). PROMs were administered remotely using REDCap. PROMs included in this analyses

were: PROM-Ataxia[31]; the depth perception, visual acuity and spatial vision, and visual processing speed sections of the Visual Activities Questionnaire[32] (VAQ); the Dysarthria Impact Scale[33] (DIS); the Communicative Participation Item Bank[34] (CPIB); and ten of the thirteen Quality of Life in Neurological Disorders (NQoL) short forms.[35] The PROM-Ataxia consists of 70 items scored on a 0–4 Likert scale distributed across five components assessing affect (MEN-1), cognition (MEN-2), activities of daily living (ADL), physical symptoms including sleep, balance, and sensation (PHYS-1), and physical capabilities (PHYS-2). The DIS (23 items) and CPIB (10 items) assess how speech impairments affect everyday life and how patients' conditions impact how they communicate on an average day, respectively. The NQoL short forms, each consisting of 8–9 items, assess the ability to perform fine motor tasks, mobility difficulties, cognition, anxiety, depression, life satisfaction, emotional control, fatigue, sleep disturbances, and patients' abilities to keep up with social responsibilities. Prior to May 16, 2023, there was a REDCap error in the first mental (affect) section of PROM-Ataxia that combined the "rarely" and "sometimes" responses (with respective scores of 2 and 3) into a single response. A value of 2.5 was used for responses in this erroneous category.

Data Processing

The time series from each eye were processed independently for every data recording session. Time series were automatically trimmed to the performance of the passage reading task using timestamps logged by the Neurobooth software (Figs. 1C and 1D). Each recording was evaluated for data quality using the methods described in the Supplemental Material. Recordings with poor quality (i.e., bad calibration, excessive head movement, or a large proportion of blinks) or a duration of less than 5 s were excluded from analysis. If the task was repeated by the participant within a single session, then only the last repetition with a duration of at least 5 s was selected.

Saccades and blinks were detected using the eye tracker's built-in algorithms. Fixations or saccades within 100 ms of a blink were excluded from analysis.[36] Detected saccades with a duration of less than 6 ms were ignored (i.e., treated as part of the adjacent fixations).[37,38] The vertical component of gaze was not considered in this analysis. Horizontal gaze angular velocities (°/s) and displacements (°) were derived as detailed in the EyeLink Portable Duo User Manual[36] and the Supplementary Material. Angular velocities were filtered using a low-pass, seventh-order Butterworth filter with a cutoff frequency of 100 Hz.[39]

Saccades with an angular displacement less than 0.5° were excluded from subsequent analysis.[38] Each saccade was then classified as a progressive saccade, regression, or sweep based on its displacement and direction. Saccades with an angular displacement of at least 15° were classified as "sweeps". This threshold was empirically determined based on the distribution of saccade angular displacements (Supplementary Fig. 1). Sweeps included both return sweeps (movements to the next line) and regressive sweeps (returning to the previous line). Non-sweep saccades were classified as regressions if participant gaze moved in a leftward direction on the screen. Otherwise, the saccade was classified as a progressive saccade.

Additional data quality filters related to the performance of the passage reading task were applied to ensure saccade-based data features could be reliability computed. More specifically, task recordings with less than forty saccades or less than four sweeps were excluded from further processing and analysis, although all task recordings with less than forty saccades also had less than four sweeps (Supplementary Fig. 2 and Supplementary Fig. 3).

Comparative Kinematic Measures

Saccade and fixation measures were computed using data from a single eye. Data from only one eye was used to compare with results reported by Terao, et al.,[25] which used monocular recordings. The eye ipsilateral to the dominant upper limb was used if available.[40–42] If the data from the ipsilateral eye was unavailable or excluded as having poor quality, then the contralateral eye was used. Kinematic measures included the mean angular displacement of saccades, the mean duration of saccades, the frequency of saccades, the mean duration of fixations, and the percentage of regressions. Saccade frequency was computed as the number of saccades divided by the cumulative duration of the saccades and fixations. In addition to these measures reported by Terao, et al., we also computed the number of saccades generated during the task recording and the mean angular displacement during fixations. (Note that the eyes are not entirely stationary during fixations.)[24] Sweeps were excluded from the calculation of all comparative kinematic measures.

Linear Model

Separate from the comparative kinematic measures, a set of 28 kinematic data features were computed to summarize each task recording. These features were extracted independently from the set of fixations and from the three saccade classes (i.e., progressive saccades, regressions, and sweeps). For fixations and each saccade class, six features were computed: the mean and coefficient of variation of duration, angular displacement, and peak angular speed. These features accounted for 24 of the 28 features. An additional feature for each saccade class was the percentage of saccades contained within the class, resulting in an additional three features. Finally, the saccade frequency feature was computed as the number of all saccades divided by the combined duration of all saccades and fixations. Supplementary Table 1 includes the complete list of features.

Lasso regression was used to jointly model the relationship between the extracted features and total BARS. The output of the regression model was therefore a machine-learned composite score of eye movement kinematics reflecting overall ataxia severity. Total BARS was used as the learning objective because the oculomotor component of BARS is coarse and non-ordinal (e.g., two scores of 1.0 can reflect a disjoint set of the four cardinal oculomotor signs), which is detrimental to the regression task. Features extracted from each eye recording were treated as independent model inputs, and the two model outputs were averaged. Though they were not scored on BARS, controls were assumed to have a total BARS of 0.0 for the purposes of model training. Only the MGH controls were included in the training data to reduce the potential for noise introduced by this assumption, although models were evaluated against all controls. Models were

trained and evaluated using leave-one-subject-out cross-validation (LOSOCV). Model outputs corresponding to the testing set of each iteration of LOSOCV were pooled to facilitate model evaluation and statistical analyses. Additional modeling details, including feature standardization and hyperparameter selection, are described in the Supplemental Material.

Statistical Analyses

The mean and standard deviation amongst controls and participants with ataxia was computed for each comparative kinematic measure. Welch's t-tests were used to determine if the mean of each measure was significantly different between controls and participants with ataxia. Effect sizes were assessed using Cohen's d. Additionally, kinematic measures for participants with ataxia were grouped into low severity, mid severity, and high severity groups corresponding to total BARS of less than 8, 8–16, and greater than or equal to 16, respectively.[43] Welch's ANOVA with Games-Howell post hoc tests were used to determine if there were significant differences in kinematic measures between controls and the three severity groups. Effect sizes for Welch's ANOVA and Games-Howell post hoc tests were assessed using partial η^2 and Cohen's d, respectively. The post-hoc tests comparing healthy controls to participants with low-severity ataxia were of particular interest.

Pearson's correlation (r) was used to assess the convergent validity of the learned composite score with total BARS, each component score of BARS, and each PROM included in the study. The ability of the composite score to distinguish between participants with ataxia and controls was assessed using a Mann-Whitney U-test. A similar test was performed for participants with ataxia and a BARS oculomotor score of 0 to assess the composite score's ability to distinguish subclinical oculomotor signs. A third U-test was used to assess whether the composite scores were different for controls recruited from the clinic and controls recruited from the general population. Effect sizes were assessed in terms of the common language effect size, which is equivalent to the area under the receiver operating characteristic curve (AUC). The correlation between participant age and the absolute error of the composite scores (with respect to total BARS) was used to determine if the accuracy of composite scores was significantly influenced by participant age. Composite score reliability was assessed by comparing the composite score for each eye during the same task recording using a two-way mixed effects, absolute agreement, single rater intraclass correlation coefficient (ICC).[44] The sensitivity of the composite score to longitudinal changes was assessed using a one-sided, one-sample t-test to determine if the mean difference in composite scores between consecutive visits was greater than zero. The effect size was assessed using Cohen's d.

All statistical analyses used an a priori significance threshold of 0.05. Interpretations of effect size strength (i.e., weak, moderate, strong) for d and partial η^2 were in accordance with guidelines proposed by Cohen.[45] Effect sizes of 0.2, 0.5, and 0.8 were interpreted as weak, moderate, and strong for Cohen's d, respectively. Effect sizes of 0.01, 0.06, and 0.14 were interpreted as weak, moderate, and strong for partial η^2, respectively.

Results

Comparative Kinematic Measures

Several saccade and fixation kinematic measures during reading were observed to differ between ataxia and control groups and change with ataxia severity (Fig. 2, Table 2). In particular, the angular displacement of detected fixations was strongly related to ataxia severity group (η^2=0.44, p<0.001). Severity was also strongly associated with an increased number of saccades (η^2=0.27, p<0.001) and reduced saccade frequency (η^2=0.25, p<0.001), and was moderately associated with an increased percentage of regressions (η^2=0.11, p<0.001). In addition to reduced saccade frequency, a moderate association was observed between severity and fixation duration (η^2=0.13, p<0.001). Notably, both the angular displacement of fixations (d=0.72, p=0.002) and the number of saccades (d=0.58, p=0.02) were able to distinguish between controls and low-severity ataxia.

Neither saccade displacement (η^2=0.03, p=0.09) nor duration (η^2=0.01, p=0.35) were found to be significantly different with respect to ataxia severity or between ataxia and control groups. We therefore hypothesized that the increased number of saccades in individuals with ataxia was due to the increased percentage of regression saccades that result in increased visual traversal of the passage. Indeed, the number of progressive and regressive saccades were correlated (r=0.55, p<0.001) and were both strongly indicative of disease severity (Supplementary Fig. 4). We also observed a strong correlation between the cumulative displacements of progressive and regressive saccades (r=0.83, p<0.001), supporting the hypothesis of increased visual traversal. It is therefore notable that the number of saccades provided a stronger signal than the percentage of regressions for both ataxia severity and for separating ataxia and control populations.

Composite Score Evaluation

Given that several saccade and fixation kinematics were strongly informative of ataxia severity, we hypothesized that aggregating kinematic features into a composite score would produce a more precise measure of severity. We trained a linear model to estimate BARS total scores based on a set of 28 kinematic features and evaluated several aspects of its performance (Fig. 3). The composite scores were both reliable (ICC=0.96, p<0.001) and had a strong correlation with total BARS (r=0.82, p<0.001). Furthermore, the scores demonstrated sensitivity to longitudinal change across multiple visits (d=0.36, p=0.03). Composite scores for participants with ataxia were not biased by age, as there was no correlation between the absolute error of composite scores (with respect to total BARS) and age (r=0.08, p=0.35). Interestingly, there was a weak correlation between age and the composite scores for healthy controls (r=0.28, p=0.001).

Next, we assessed the ability of the composite scores to separate healthy control and ataxia populations. Composite scores accurately distinguished between controls and participants with ataxia (AUC=0.89, p<0.001). There was no difference in the composite scores of general population controls and MGH controls (AUC=0.52, p=0.76). Composite scores also showed sensitivity for capturing subtle early signs of ataxia: a significant difference was observed between

the scores assigned to controls and participants with ataxia that did not have clinically-observable oculomotor signs (AUC=0.69, p=0.02) (Fig. 4A).

We further investigated the relationship between composite scores and the BARS oculomotor and speech subscores (Fig. 4). Composite scores were moderately correlated with BARS oculomotor (r=0.52, p<0.001) and strongly correlated with BARS speech (r=0.73, p<0.001). We hypothesized that the stronger correlation with BARS speech was due to a stronger relationship between BARS speech and overall ataxia severity compared to the BARS oculomotor subscore. To test this hypothesis, we computed the correlation between each BARS subscore and the remainder of BARS. Indeed, BARS oculomotor was moderately correlated with the remainder of BARS (r=0.50, p<0.001) and BARS speech was strongly correlated with the remainder of BARS (r=0.77, p<0.001).

As the study population includes a diverse range of ataxias, we then evaluated the relationship between composite scores and BARS total separately for the six most-frequent ataxia diagnoses in the study (Fig. 5). The four most-frequent diagnoses (SCA-3, FRDA, SCA-6, and CANVAS) all demonstrated similar trends to the composite score across all diagnoses. The composite scores for MSA-C and SCA-2 exhibited some differences with respect to overall composite score. The sample size was smallest for MSA-C and SCA-2 (N=5 and N=6, respectively) and participants spanned a narrow range of disease severity (Fig. 5).

Finally, we investigated relationships between the composite score, based on eye movements during passage reading, and self-reported function across a wide range of PROM instruments (Table 3). Composite scores were moderately correlated with PROMs that captured motor symptoms of ataxia, including the PROM-Ataxia total score, the PHYS-2 (physical capability) and ADL (activities of daily living) components of PROM-Ataxia, and the NQoL upper and lower extremity short forms (fine motor capabilities and mobility, respectively). As an exception, the PHYS-1 component of PROM-Ataxia was only weakly correlated with composite scores. In addition to motor questions, PHYS-1 contains questions relating to autonomic, sensory, and sleep dysfunction. Composite scores were also moderately correlated with PROMs capturing speech symptoms and dysarthria, including the Communicative Participation Item Bank and Dysarthria Impact Scale. Though there was a weak correlation with visual processing speed as measured by VAQ, there was no significant correlation with depth perception or visual acuity and spatial vision. There were also no significant correlations with PROMs capturing mental and emotional state, psychiatric factors, and sleep (i.e., the MEN-1 component of PROM-Ataxia and the NQoL fatigue, sleep, depression, anxiety, and emotional dyscontrol short forms). Finally, although there was a weak correlation with the MEN-2 (cognition) component of PROM-Ataxia, there was no significant correlation with the NQoL cognition short form.

Feature Importance

The severity estimation model was probed to understand the relative contribution of each saccade and fixation feature to the learned composite score. Higher composite scores (indicating increased disease severity) were predominantly driven by a decreased percentage of sweeps,

increased fixation displacement, decreased saccade frequency, increased variability of fixation durations, and an increased percentage of regressions (Supplementary Table 1). These feature importances are consistent with the strong trends observed in the kinematic measures of saccades and fixations (Fig. 2 and Table 2). To evaluate the necessity of computing features separately for progressive saccades, regressions, and sweeps, we ran an experiment in which there was no distinction made between different types of saccades (Supplemental Material). All composite score performance metrics were reduced as a result of this change.

Discussion

We have shown that video-oculography-based analysis of eye movements during passage reading can produce accurate, reliable, and sensitive measures of ataxia. In particular, the mean displacement of fixations, the number and frequency of saccades, and the percentage of regressions discriminated between ataxia and control participants and reflected different levels of disease severity. Aggregating these saccade and fixation features using an interpretable, machine-learned composite severity score showed high reliability, responsiveness to disease progression, and sensitivity for capturing signs of early disease. These measures were derived from a brief assessment consisting only of a five-point calibration and the reading of ten lines of text.

Prior studies support our findings that particular kinematic properties of eye movement are informative of the presence and severity of ataxia. The mean displacement of fixations captures fixational instability, which is a hallmark of the ataxia phenotype.[46–49] A prior study by Oh, et al. also observed greater fixation dispersion during regularly-spaced number reading.[23] The increased percentage of regression saccades in ataxia participants was previously observed when reading Japanese text aloud.[25] The increased percentage of regressions likely contributed to both the increased number of progressive saccades[23] and reduced percentage of saccades that were sweep saccades (due to the increased number of regressive and progressive saccades per line). The increased percentage of regressions and larger number of total saccades may contribute to additional concentration and fatigue during reading and other tasks requiring sequential control of eye movements.[23] We also observed a strong negative relationship between saccade frequency (i.e., saccades per second) and ataxia severity that was not observed by Terao, et al.[25] Our identification of this signal is likely due to the larger size of our cohort and the use of a more fine-grained proxy for severity (i.e., BARS) than disease stage. Additional factors, including the language of the reading task (which is logographic and has a significant vertical component) may also have contributed to this finding. A reduction in saccade frequency is also supported by prior studies of silent reading that reported slower reading speeds[23,49] and increased fixation duration.[23] Another interesting finding is that we did not observe significant differences in the displacement, peak speed, or duration of progressive saccades. The lack of difference in mean saccade displacements may be due to the small and variable size of saccades generated during natural performance of the passage reading task. The lack of difference in mean saccade durations or peak speeds may additionally be due to the heterogeneity of the diagnoses included in this study, as not all ataxias are associated with saccadic slowing. Finally, the relative importance of features in our trained model and our experimental removal of features specific to

each saccade type convey the importance of characterizing the kinematics of sweeps in addition to the kinematics of progressive and regressive saccades.[23]

The learned composite score more strongly correlated with clinician-assessed and patient-reported measures of overall function and speech function compared to visual function. This was an interesting observation given that the composite score was based on eye movement kinematic data alone. This observation may be due in part to the fact that the composite score was trained to estimate BARS total rather than BARS oculomotor. This choice was made because the BARS oculomotor score, which counts the number of oculomotor signs, is not an ordinal variable and lacks granularity in comparison with BARS total. However, it is probable that eye movements during reading capture different aspects of disease, beyond the four cerebellar oculomotor signs assessed in the BARS oculomotor subscore. For example, dysarthria and consequent slower reading has been shown in ataxia populations to result in a larger percentage of regressions and reduced saccade frequency because eye movements are slowed to not outpace speech production.[24,25] In contrast, the increased displacement of fixations is likely influenced by nystagmus and fixational instability, which are considered by BARS oculomotor. Therefore, the composite score may better reflect speech and overall function because it includes information from more than one motor domain. Similarly, the reading task is not well-suited to characterize smooth pursuit deficits and the ability to characterize dysmetric saccades is limited by the small displacement of progressive and regressive saccades—though future work could more closely consider undersweep and oversweep during line transitions. The inability to model these oculomotor signs likely contributed to the weaker correlation between the composite score and BARS oculomotor. In addition to different information captured by quantitative eye tracking data during passage reading and clinician oculomotor assessments, the weak correlation between the composite score and BARS oculomotor score may be in part due to limited precision of the human-rated assessment. The effect of limited precision could be assessed in future work, through analysis of video-oculography-based assessments of eye movement kinematics during pure oculomotor tasks (e.g., horizontal saccades).[14]

It is also interesting that the composite score did not reflect patient-reported visual function as measured by the Visual Activities Questionnaire. Though some questions on VAQ ask about reading, they are designed to relate to factors such as visual acuity (e.g., trouble reading small print) or visual processing speed (e.g., trouble reading moving text) as opposed to the difficulty of reading itself. We did, however, observe a weak correlation between visual processing speed and composite scores, which may reflect the increased fixation duration, regression percentage, and number of saccades of ataxia participants during the reading task. In contrast, the visual acuity and depth perception components of VAQ may be more influenced by ocular motility defects[50] that were not well-captured by the reading task. More specifically, the large font and high contrast of the passage, controlled lighting in the booth, and stationary position of the monitor may have reduced the impact of impaired visual acuity, depth perception, and vergence. Future work varying these controlled factors may lead to measures that more closely reflect these particular aspects of visual function.

The ability to quantitatively assess eye movements during reading in the clinic motivates the possibility of more ecologically valid assessments of eye movements during natural reading at home. As reading is fundamental to many aspects of everyday modern life, it is an ideal target for a common, meaningful, and structured behavior to capture and assess in the real world. Recent advances in the use of mobile phone cameras for eye tracking,[51–54] in conjunction with innovations in smart glasses and extended reality technologies, may soon provide the technical capability to enable high quality recording of eye movements at home and at scale. Because the composite score presented in this work only requires knowledge of the duration, speed, and relative displacement (as opposed to absolute position) of eye movements, the results may generalize to eye movements recorded while reading text using these devices. However, additional research is needed to validate the usability of kinematic measures obtained from these devices in realistic settings given the potential impacts of varied environmental conditions and slippage[55] on data quality. Despite reduced data quality with respect to research-grade eye trackers, as commercial eye tracking technologies become more commonplace researchers are likely to curate larger and more comprehensive data sets of eye movement recordings to support algorithmic development. Reading is likely to be included in these datasets as it is a common, functionally relevant behavior, and has already been included in existing data sets of normative eye movements.[56,57] Digital assessments could leverage this abundance of normative data to support more precise measures. For example, models for detecting early signs could condition their output on a range of demographic, genetic, and visual health factors. The multitude of avenues for the application of reading-based assessments merits additional investigation across a range of other cognitive and movement disorders.[58,59] For example, eye movement abnormalities occur early in the course of Huntington's disease[58,60] and are indicative of progression.[61] Further work should therefore investigate whether methods similar to those detailed herein can be used to identify reading-based digital biomarkers of eye movement abnormalities across a broader spectrum of diseases, which may complement existing wearable[62] or in-clinic[18] digital biomarkers derived from other motor functions.

There were some limitations to the study. Data were collected as part of a larger study in a clinic setting under time constraints, which may have contributed to the need to exclude some sessions due to poor calibration. These exclusions resulted in a slight inclusion bias towards individuals with milder disease severity, although more severe individuals were well represented in the study. Including real time feedback on data quality and allowing additional time for setup and calibration (and repetition of the task when needed) could substantially reduce the percentage of sessions with poor data quality. Additional time would also permit reading of multiple passages that would provide additional data to further reduce the likelihood of exclusion. It is also possible that some controls included in the study were not true controls, as we relied on self-reported medical history rather than a detailed neurological assessment of controls. However, the learned composite score was able to distinguish between controls and participants with ataxia despite this possibility. Only the horizontal component of gaze was analyzed in this study. Reading in other languages has a more significant vertical component, which may complicate study replication in many geographic regions.

Video oculography has clear potential to power precise and objective digital assessments of eye movements during reading in ataxias. The in-clinic data collection protocol using the Neurobooth platform demonstrates the scalability and practicality of such assessments. We anticipate that ongoing data collection will enable the development of more sophisticated models with increased sensitivity to detect subtle early signs and measure disease progression. For example, computational models of eye movements sequences (e.g., using transformer neural networks[63]) will likely be able to leverage temporal information not captured by the kinematic measures reported herein. These results support the integration of quantitative oculomotor assessments of reading into routine clinical practice[23] and clinical research in ataxias.

Data Availability

Data can be requested by qualified researchers by visiting https://neurobooth.mgh.harvard.edu/.

Acknowledgements

This project utilized resources at the Martinos-MLSC cluster. The authors would like to thank prior contributors to the Neurobooth project not directly involved with this research, including Adonay Nunes, Mainak Jas, Nicole Eklund, and Sheraz Khan. The authors also thank the Neurobooth participants for their valuable time and perspectives.

Funding

Massachusetts Life Sciences Center, NIH (NS117826), Biogen, Broad Institute, Dake Family Foundation, Massachusetts General Hospital Department of Neurology.

Competing Interests

Dr. Jeremy Schmahmann is the inventor of the Brief Ataxia Rating Scale and the Patient Reported Outcome Measure of Ataxia (PROM-Ataxia), which are licensed to the General Hospital Corporation and were used in this study.

References

1. Joseph Jankovic, Mark Hallett, Michael S. Okun, Cynthia Comella, Stanley Fahn, Jennifer Goldman. Ataxia: Pathophysiology and clinical syndromes. In: *Principles and Practice of Movement Disorders*. Elsevier; 2021:524-536.

2. Ashizawa T, Xia G. Ataxia. *Continuum*. 2016;22(4 Movement Disorders):1208-1226.

3. López-Bastida J, Perestelo-Pérez L, Montón-Alvarez F, Serrano-Aguilar P. Social economic costs and health-related quality of life in patients with degenerative cerebellar ataxia in

Spain. *Mov Disord*. 2008;23(2):212-217.

4. Wilson CL, Fahey MC, Corben LA, et al. Quality of life in Friedreich ataxia: what clinical, social and demographic factors are important? *Eur J Neurol*. 2007;14(9):1040-1047.

5. Stephen CD, Schmahmann JD. Eye movement abnormalities are ubiquitous in the spinocerebellar ataxias. *Cerebellum*. 2019;18(6):1130-1136.

6. Moscovich M, Okun MS, Favilla C, et al. Clinical evaluation of eye movements in spinocerebellar ataxias: a prospective multicenter study. *J Neuroophthalmol*. 2015;35(1):16-21.

7. Kheradmand A, Zee DS. Cerebellum and ocular motor control. *Front Neurol*. 2011;2:53.

8. Paulson HL, Shakkottai VG, Clark HB, Orr HT. Polyglutamine spinocerebellar ataxias - from genes to potential treatments. *Nat Rev Neurosci*. 2017;18(10):613-626.

9. Scoles DR, Pulst SM. Oligonucleotide therapeutics in neurodegenerative diseases. *RNA Biol*. 2018;15(6):707-714.

10. Ghanekar SD, Kuo SH, Staffetti JS, Zesiewicz TA. Current and emerging treatment modalities for spinocerebellar ataxias. *Expert Rev Neurother*. 2022;22(2):101-114.

11. Schmahmann JD, Gardner R, MacMore J, Vangel MG. Development of a brief ataxia rating scale (BARS) based on a modified form of the ICARS. *Mov Disord*. 2009;24(12):1820-1828.

12. Schmitz-Hübsch T, du Montcel ST, Baliko L, et al. Scale for the assessment and rating of ataxia: development of a new clinical scale. *Neurology*. 2006;66(11):1717-1720.

13. Trouillas P, Takayanagi T, Hallett M, et al. International Cooperative Ataxia Rating Scale for pharmacological assessment of the cerebellar syndrome. The Ataxia Neuropharmacology Committee of the World Federation of Neurology. *J Neurol Sci*. 1997;145(2):205-211.

14. Garces P, Antoniades CA, Sobanska A, et al. Quantitative oculomotor assessment in hereditary ataxia: Systematic review and consensus by the ataxia global initiative working group on digital-motor biomarkers. *Cerebellum*. 2024;23(3):896-911.

15. Gupta AS. Digital Phenotyping in Clinical Neurology. *Semin Neurol*. 2022;42(1):48-59.

16. Torri F, Vadi G, Meli A, et al. The use of digital tools in rare neurological diseases towards a new care model: a narrative review. *Neurol Sci*. Published online June 10, 2024. doi:10.1007/s10072-024-07631-4

17. Poleur M, Markati T, Servais L. The use of digital outcome measures in clinical trials in rare neurological diseases: a systematic literature review. *Orphanet J Rare Dis*. 2023;18(1):224.

18. Reilmann R, Schubert R. Motor outcome measures in Huntington disease clinical trials. *Handb Clin Neurol*. 2017;144:209-225.

19. Sekar A, Panouillères MTN, Kaski D. Detecting Abnormal Eye Movements in Patients with Neurodegenerative Diseases - Current Insights. *Eye Brain*. 2024;16:3-16.

20. Elliott DB, Trukolo-Ilic M, Strong JG, Pace R, Plotkin A, Bevers P. Demographic characteristics of the vision-disabled elderly. *Invest Ophthalmol Vis Sci*. 1997;38(12):2566-2575.

21. Rubin GS. Measuring reading performance. *Vision Res*. 2013;90:43-51.

22. Brown JC, Goldstein JE, Chan TL, Massof R, Ramulu P, Low Vision Research Network Study Group. Characterizing functional complaints in patients seeking outpatient low-vision services in the United States. *Ophthalmology*. 2014;121(8):1655-1662.e1.

23. Oh AJ, Chen T, Shariati MA, Jehangir N, Hwang TN, Liao YJ. A simple saccadic reading test to assess ocular motor function in cerebellar ataxia. *PLoS One*. 2018;13(11):e0203924.

24. Rayner K. Eye movements in reading and information processing: 20 years of research. *Psychol Bull*. 1998;124(3):372-422.

25. Terao Y, Tokushige SI, Inomata-Terada S, et al. How do patients with Parkinson's disease and cerebellar ataxia read aloud? -Eye-voice coordination in text reading. *Front Neurosci*. 2023;17:1202404.

26. Nunes AS, Patel S, Oubre B, et al. Multimodal Digital Phenotyping of Behavior in a Neurology Clinic: Development of the Neurobooth Platform and the First Two Years of Data Collection. *medRxiv*. Published online January 1, 2024:2024.12.28.24319527.

27. Barnett C, Green JR, Marzouqah R, et al. Reliability and validity of speech & pause measures during passage reading in ALS. *Amyotroph Lateral Scler Frontotemporal Degener*. 2020;21(1-2):42-50.

28. Olsen L, DePalma L, Evans JH. Self-interested and altruistic motivations in volunteering for clinical trials: A more complex relationship. *J Empir Res Hum Res Ethics*. 2020;15(5):443-451.

29. Stunkel L, Grady C. More than the money: a review of the literature examining healthy volunteer motivations. *Contemp Clin Trials*. 2011;32(3):342-352.

30. McCann SK, Campbell MK, Entwistle VA. Reasons for participating in randomised controlled trials: conditional altruism and considerations for self. *Trials*. 2010;11(1):31.

31. Schmahmann JD, Pierce S, MacMore J, L'Italien GJ. Development and Validation of a Patient-Reported Outcome Measure of Ataxia. *Mov Disord*. 2021;36(10):2367-2377.

32. Sloane M, Ball K, Owsley C, Bruni J, Roenker D. The visual activities questionnaire: Developing an instrument for assessing problems in everyday visual tasks. *Noninvasive Assessment of the Visual System*. Published online 1992. doi:10.1364/navs.1992.sub4

33. Vogel A, Graf L, Chan J, Hepworth G, Bade M, Synofzik M. Development and validation of the dysarthria impact scale. Presented at: International Congress on Ataxia Research; 2022.

34. Baylor C, Yorkston K, Eadie T, Kim J, Chung H, Amtmann D. The Communicative Participation Item Bank (CPIB): item bank calibration and development of a disorder-generic short form. *J Speech Lang Hear Res*. 2013;56(4):1190-1208.

35. Cella D, Lai JS, Nowinski CJ, et al. Neuro-QOL: brief measures of health-related quality of life for clinical research in neurology. *Neurology*. 2012;78(23):1860-1867.

36. SR-Research, Ltd. *EyeLink® Portable Duo User Manual, Version 1.0.9*. Accessed July 17, 2024. https://www.sr-research.com/support/thread-174.html

37. Engbert R, Mergenthaler K. Microsaccades are triggered by low retinal image slip. *Proc Natl Acad Sci U S A*. 2006;103(18):7192-7197.

38. Rayner K. Eye movement latencies for parafoveally presented words. *Bull Psychon Soc*. 1978;11(1):13-16.

39. Raju MH, Friedman L, Bouman TM, Komogortsev OV. Filtering Eye-Tracking Data From an EyeLink 1000: Comparing Heuristic, Savitzky-Golay, IIR and FIR Digital Filters. *arXiv [csHC]*. Published online March 3, 2023. http://arxiv.org/abs/2303.02134

40. McManus IC, Porac C, Bryden MP, Boucher R. Eye-dominance, writing hand, and throwing hand. *Laterality*. 1999;4(2):173-192.

41. Zeri F, De Luca M, Spinelli D, Zoccolotti P. Ocular dominance stability and reading skill: a controversial relationship. *Optom Vis Sci*. 2011;88(11):1353-1362.

42. Miller PM. *Binocular and Monocular Assessment of Eye Movements – Does Dominance Confer a Performance Advantage?* The Ohio State University; 2024. Accessed January 9, 2025. https://etd.ohiolink.edu/acprod/odb_etd/ws/send_file/send?accession=osu1713504180952517&disposition=inline

43. Oubre B, Daneault JF, Whritenour K, et al. Decomposition of Reaching Movements Enables Detection and Measurement of Ataxia. *Cerebellum*. Published online March 2, 2021. doi:10.1007/s12311-021-01247-6

44. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-163.

45. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Routledge; 2013. doi:10.4324/9780203771587

46. Hocking DR, Fielding J, Corben LA, et al. Ocular motor fixation deficits in Friedreich ataxia. *Cerebellum*. 2010;9(3):411-418.

47. Kim JS, Kim JS, Youn J, et al. Ocular motor characteristics of different subtypes of spinocerebellar ataxia: distinguishing features: Ocular Motor Features of Sca Subtypes. *Mov Disord*. 2013;28(9):1271-1277.

48. Federighi P, Ramat S, Rosini F, Pretegiani E, Federico A, Rufa A. Characteristic Eye Movements in Ataxia-Telangiectasia-Like Disorder: An Explanatory Hypothesis. *Front Neurol*. 2017;8:596.

49. Alexandre MF, Rivaud-Péchoux S, Challe G, Durr A, Gaymard B. Functional Consequences of Oculomotor Disorders in Hereditary Cerebellar Ataxias. *Cerebellum*. 2013;12(3):396-405.

50. Kedar S, Ghate D, Murray EL, Corbett JJ, Subramony SH. Vision related quality of life in spinocerebellar ataxia. *J Neurol Sci*. 2015;358(1-2):404-408.

51. Barahim Bastani P, Saber Tehrani AS, Badihian S, et al. Self-recording of eye movements in amyotrophic lateral sclerosis patients using a smartphone eye-tracking app. *Digit Biomark*. 2024;8(1):111-119.

52. Németh AH, Antoniades CA, Dukart J, et al. Using Smartphone Sensors for Ataxia Trials: Consensus Guidance by the Ataxia Global Initiative Working Group on Digital-Motor Biomarkers. *Cerebellum*. Published online November 28, 2023. doi:10.1007/s12311-023-01608-3

53. Azami H, Chang Z, Arnold SE, Sapiro G, Gupta AS. Detection of Oculomotor Dysmetria From Mobile Phone Video of the Horizontal Saccades Task Using Signal Processing and Machine Learning Approaches. *IEEE Access*. 2022;10:34022-34031.

54. Chang Z, Chen Z, Stephen CD, et al. Accurate detection of cerebellar smooth pursuit eye movement abnormalities via mobile phone video and machine learning. *Sci Rep*. 2020;10(1):18641.

55. Niehorster DC, Santini T, Hessels RS, Hooge ITC, Kasneci E, Nyström M. The impact of slippage on the data quality of head-worn eye trackers. *Behav Res Methods*. 2020;52(3):1140-1160.

56. Griffith H, Lohr D, Abdulin E, Komogortsev O. GazeBase, a large-scale, multi-stimulus, longitudinal eye movement dataset. *Sci Data*. 2021;8(1):184.

57. Bolliger LS, Haller P, Cretton ICR, Reich DR, Kew T, Jäger LA. EMTeC: A corpus of Eye Movements on machine-generated texts. *arXiv [csCL]*. Published online August 8, 2024. http://arxiv.org/abs/2408.04289

58. Anderson TJ, MacAskill MR. Eye movements in patients with neurodegenerative disorders. *Nat Rev Neurol*. 2013;9(2):74-85.

59. White OB, Fielding J. Cognition and eye movements: assessment of cerebral dysfunction. *J Neuroophthalmol*. 2012;32(3):266-273.

60. Grabska N, Wójcik-Pędziwiatr M, Sławek J, et al. Reflexive and voluntary saccadic eye movements as biomarker of Huntington's Disease. *Neurol Neurochir Pol*. Published online May 31, 2024. doi:10.5603/pjnns.95190

61. Hicks SL, Robert MPA, Golding CVP, Tabrizi SJ, Kennard C. Oculomotor deficits indicate the progression of Huntington's disease. Kennard C, Leigh RJ, eds. *Prog Brain Res*. 2008;171:555-558.

62. Gupta AS, Patel S, Premasiri A, Vieira F. At-home wearables and machine learning sensitively capture disease progression in amyotrophic lateral sclerosis. *Nat Commun*. 2023;14(1):5080.

63. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *arXiv [csCL]*. Published online June 12, 2017. Accessed February 3, 2023. https://proceedings.neurips.cc/paper/7181-attention-is-all

Figures

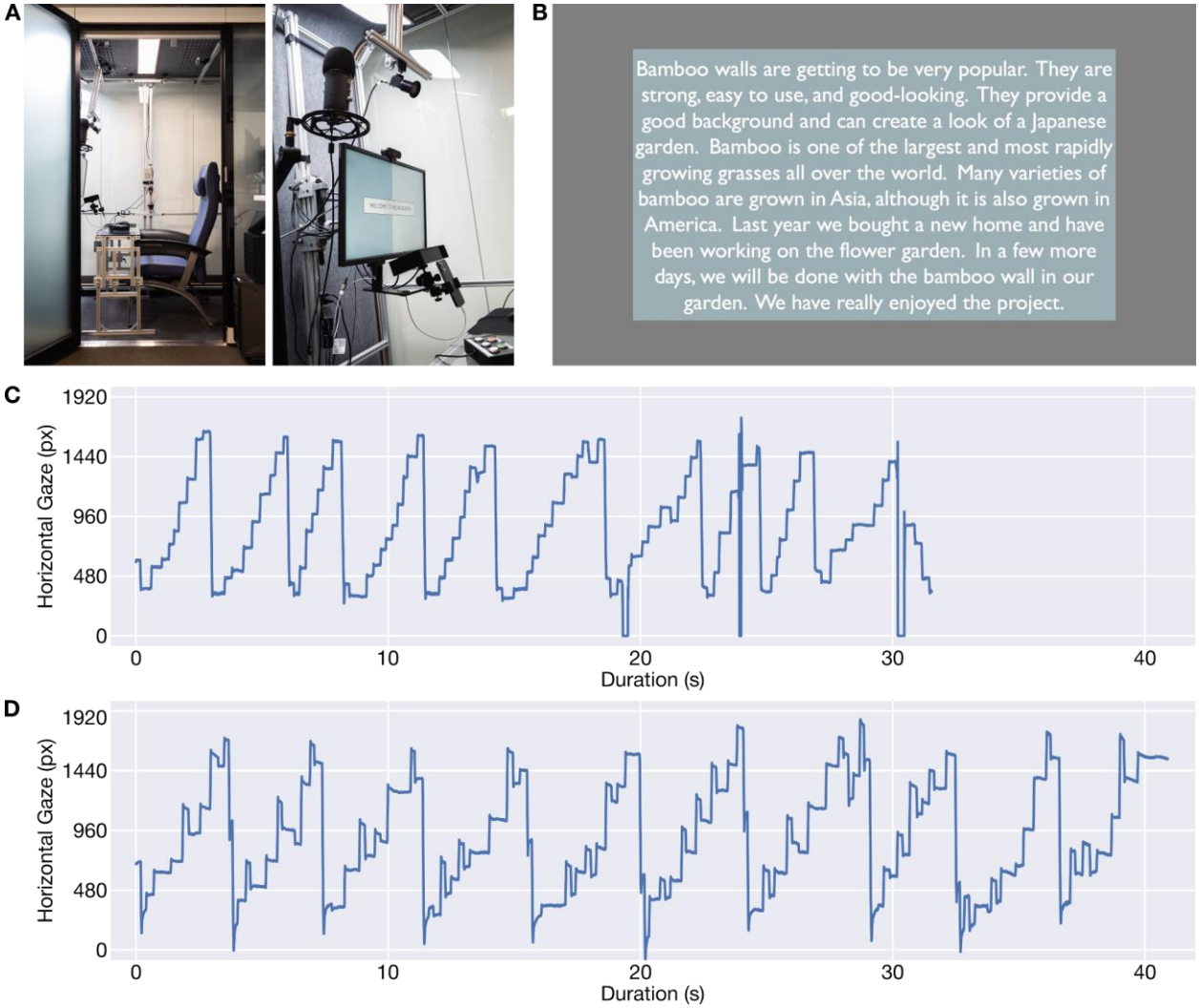


Figure 1. Experimental setup and data visualization. (**A**) The physical configuration of the Neurobooth, including the relative position of the chair, monitor, and EyeLink Portable Duo. (**B**) The Bamboo passage[27] presented on the monitor to participants during the passage reading task. (**C**) The horizontal component of gaze during passage reading for a healthy control. (**D**) The horizontal component of gaze during passage reading for a participant with ataxia. The participant with ataxia took longer to read the passage and had more frequent regressions. Nystagmus, dysmetric saccades, saccadic intrusions, and saccadic pursuit were observed during clinical examination.

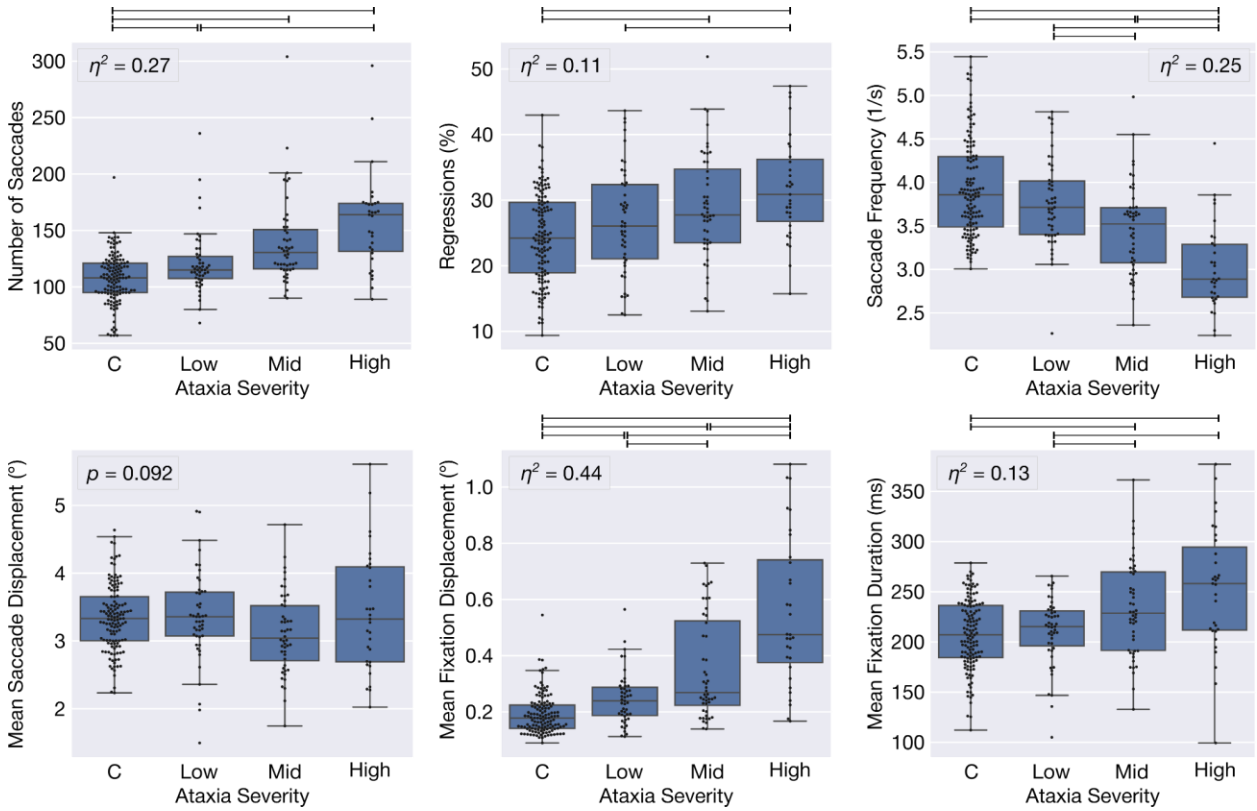


Figure 2. Comparative kinematic measures of interest. Sweeps were excluded from measure calculations. Participants were grouped into controls (C), low severity, mid severity, and high severity based on the total Brief Ataxia Rating Scale. Box plots denote each quartile, with a maximum whisker length of 1.5 times the interquartile range. The overlaid swarm shows each task recording. The horizontal black bars above each plot denote statistical significance of the post-hoc tests comparing the indicated groups. The omnibus test was not significant for mean saccade displacement.

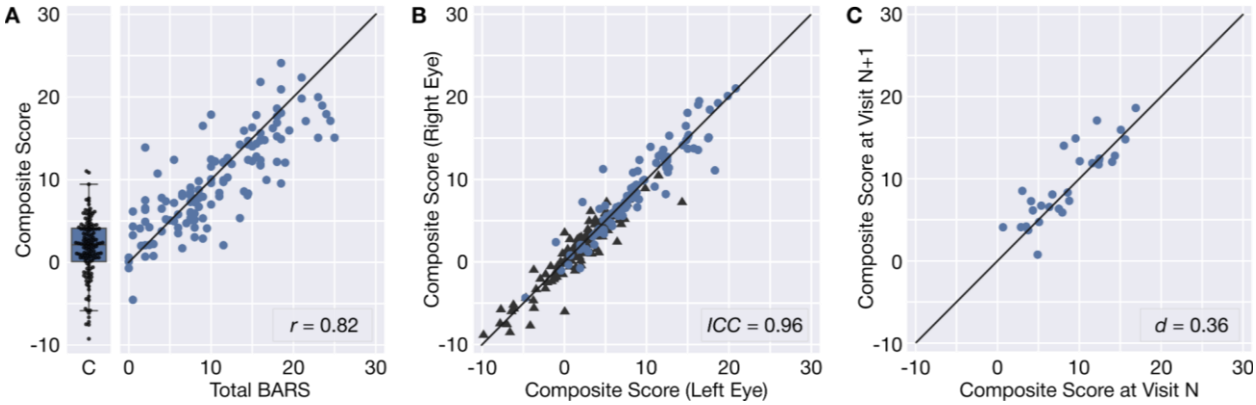


Figure 3. Validation and characterization of per-recording learned composite scores. In all plots, the solid black line denotes perfect agreement and markers correspond to a single task recording. (**A**) Composite scores (averaged across both eyes) strongly correlate with total Brief Ataxia Rating Scale (BARS) scores. The distribution of composite scores is shown separately as a swarm for healthy controls, as BARS was not assessed. The box plot denotes each quartile, with a maximum whisker length of 1.5 times the interquartile range. (**B**) Composite scores independently obtained from each eye from the same task recording demonstrate very strong agreement. Healthy controls are denoted by black triangles. (**C**) Composite scores from consecutive visits (at least 85 days apart) demonstrate sensitivity to longitudinal change; validated by a one-sided, one-sample *t*-test.

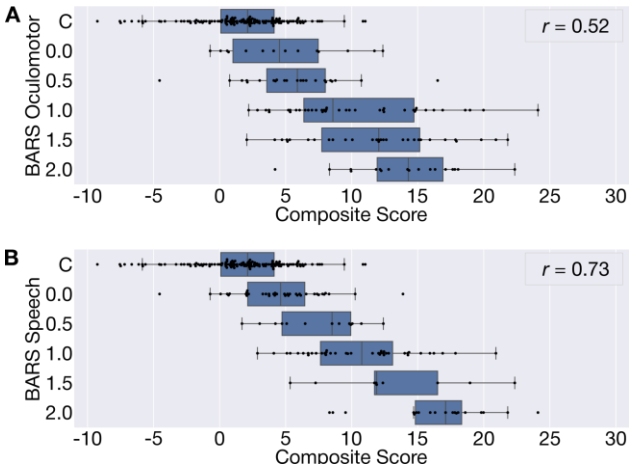


Figure 4. Composite scores versus (A) BARS Oculomotor and (B) BARS Speech. Composite scores increase with respect to BARS subscores and demonstrate some sensitivity to subclinical signs. No participant with ataxia had a BARS Speech score of more than 2. Box plots denote each quartile, with a maximum whisker length of 1.5 times the interquartile range. The overlaid swarm shows each task recording.

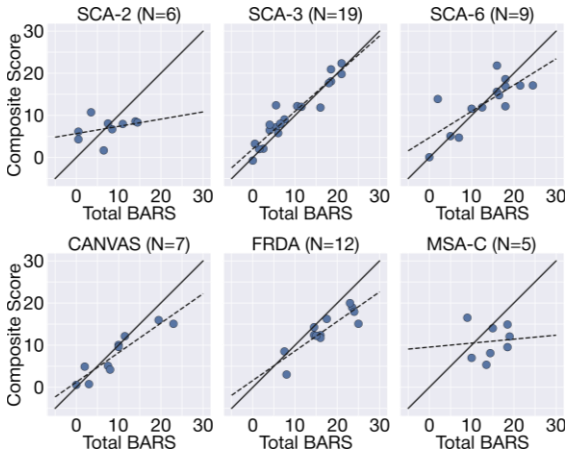


Figure 5. Per-recording composite scores versus Brief Ataxia Rating Scale (BARS) scores for the six most frequent ataxia diagnoses in the data set. Each point represents a single task recording (participants could have multiple task recordings from different timepoints). The solid black line denotes perfect agreement with total Brief Ataxia Rating Scale (BARS) scores and the dotted black line denotes a least-squares regression.

Abbreviations: SCA—Spinocerebellar ataxia; CANVAS—Cerebellar Ataxia, Neuropathy, and Vestibular Areflexia Syndrome; FRDA—Friedreich's Ataxia; MSA-C—Multiple System Atrophy (Cerebellar Type)

Tables

Table 1. Participant demographics. “Other Ataxias” included Spinocerebellar Ataxia Types 4, 7, 9, 17, 19, 22, 28, and 43; Gordon-Holmes syndrome; Behcet’s disease; autosomal recessive cerebellar ataxia; sensory ataxia; vestibular ataxia; pancerebellar syndrome; autoimmune-related ataxia with undefined cause; sporadic adult onset ataxia; and sporadic adult onset ataxia with neuropathy; and sporadic adult onset ataxia with downbeat nystagmus.

Abbreviations: BARS—Brief Ataxia Rating Scale; MSA-C—Multiple System Atrophy (Cerebellar Type); CANVAS—Cerebellar Ataxia, Neuropathy, and Vestibular Areflexia Syndrome; HSP-7—Hereditary Spastic Paraplegia Type 7

Diagnosis	N	Pct. Female	Decade of Life	BARS Total (Mean, Range)	BARS Oculomotor (Mean, Range)
Control	**70**	**64**	**2nd–8th**	—	—
Total Ataxia	**102**	**47**	**2nd–8th**	10.6 (0.0–25.0)	1.08 (0.0–2.0)
Spinocerebellar Ataxia Type 1	2	100	3rd–6th	13.5 (8.5–18.5)	0.75 (0.5–1.0)
Spinocerebellar Ataxia Type 2	6	50	3rd–7th	7.3 (0.5–14.0)	0.83 (0.5–1.5)
Spinocerebellar Ataxia Type 3	19	63	2nd–7th	9.3 (0.0–21.0)	1.14 (0.0–2.0)
Spinocerebellar Ataxia Type 6	9	56	6th–8th	12.2 (0.0–24.5)	1.22 (0.0–2.0)
Friedreich's Ataxia	12	33	2nd–7th	18.1 (8.0–25.0)	1.15 (0.0–2.0)
MSA-C	5	60	6th–7th	13.1 (9.0–18.5)	1.00 (0.5–1.5)
CANVAS	7	29	3rd–7th	8.6 (0.0–23.0)	1.21 (0.0–2.0)
HSP-7	4	25	4th–6th	12.9 (7.5–16.0)	0.88 (0.0–1.5)
Other Ataxias	38	42	2nd–8th	8.8 (0.5–18.0)	1.07 (0.0–2.0)

Table 2. Kinematic Measures of Saccades and Fixations. Measures were computed for each task recording. The mean and standard deviation were computed with respect to all sessions within each population. The comparisons in parentheses are originally reported in Table 3 of Terao, et. al[25], and correspond to reading Japanese texts aloud. Effect size statistics (Cohen's d) compared controls and participants with ataxia in this study.

Measure	Controls (Comparison)	Ataxia (Comparison)	Effect Size
Number of saccades	107.3 ± 22.1 (—)	131.3 ± 38.0 (—)	d=0.96, p<0.001
Mean saccade displacement (°)	3.35 ± 0.49 (3.88 ± 1.62)	3.30 ± 0.72 (3.56 ± 1.14)	d=0.08, p=0.51
Mean saccade duration (ms)	41.3 ± 13.7 (74.3 ± 38.9)	42.3 ± 24.3 (57.3 ± 30.8)	d=0.05, p=0.67
Mean saccade peak speed (°/s)	279.4 ± 122.5 (—)	254.7 ± 108.3 (—)	d=0.21, p=0.08
Saccade frequency (1/s)	3.93 ± 0.56 (5.06 ± 1.21)	3.47 ± 0.59 (3.80 ± 0.71)	d=0.80, p<0.001
Regression (%)	24.4 ± 6.7 (22.9 ± 6.6)	28.8 ± 8.2 (27.1 ± 8.5)	d=0.58, p<0.001
Fixation displacement (°)	0.19 ± 0.07 (—)	0.36 ± 0.21 (—)	d=1.06, p<0.001
Fixation duration (ms)	207.7 ± 34.9 (236.8 ± 70.4)	230.5 ± 49.9 (293.3 ± 60.9)	d=0.53, p<0.001

Table 3. Correlations of contextual clinical measures with learned composite scores.

Abbreviations: PROM—Patient-Reported Outcome Measure; VAQ—Visual Activities Questionnaire; DIS—Dysarthria Impact Scale; CPIB—Communicative Participation Item Bank; NQoL—Quality of Life in Neurological Disorders short forms

PROM	PROM Component	Correlation with Composite Score
PROM-Ataxia	(total)	r=0.51, p<0.001
	PHYS-1	r=0.37, p<0.001
	PHYS-2	r=0.62, p<0.001
	ADL	r=0.56, p<0.001
	MEN-1	r=0.14, p=0.13
	MEN-2	r=0.35, p<0.001
VAQ	Depth Perception	r=0.17, p=0.10
	Visual Processing Speed	r=0.27, p=0.01
	Visual Acuity and Spatial Vision	r=0.00, p=0.99
DIS	(total)	r=0.53, p<0.001
CPIB	(total)	r=-0.54, p<0.001
NQoL	Lower Extremity	r=-0.53, p<0.001
	Upper Extremity	r=-0.47, p<0.001
	Participation in Social Roles	r=-0.34, p<0.001
	Wellbeing	r=-0.22, p=0.02
	Cognition	r=-0.13, p=0.17

PROM	PROM Component	Correlation with Composite Score
PROM-Ataxia	(total)	r=0.51, p<0.001
	PHYS-1	r=0.37, p<0.001
	PHYS-2	r=0.62, p<0.001
	ADL	r=0.56, p<0.001
	MEN-1	r=0.14, p=0.13
	MEN-2	r=0.35, p<0.001
	Fatigue	r=-0.01, p=0.95
	Sleep	r=0.00, p=0.97
	Depression	r=0.13, p=0.18
	Anxiety	r=0.05, p=0.57
	Emotional Dyscontrol	r=0.15, p=0.12

Supplementary Material

Exclusion of Poor-Quality Data

The quality of each task recording was assessed using three criteria. The first criteria was designed to reject recordings with poor calibration. Every sample where the horizontal or vertical gaze position exceeded the known boundaries of the screen by at least 20% were flagged as out-of-bounds. The second criteria was designed to reject recordings with excessive missing data. Every sample that had a zero (i.e., placeholder) value or that was labeled as a blink by the eye tracker was flagged as being missing data. If at least 20% of samples in the task recording were flagged as either out-of-bounds or missing, then the task recording was excluded from analysis. The third criteria was designed to reject recordings where the participant's head excessively moved. If the standard deviation of the distance between the forehead marker and the eye tracker during the recording exceeded 100 mm, then the recording was excluded from analysis. After data quality exclusions, the mean, 95th percentile, and maximum of the standard deviation of the forehead marker distance in each recording was 3.6 mm, 10.7 mm, and 31.1 mm, respectively.

Excluded Participants

The BARS total score for excluded participants with ataxia ranged from 2.5–23.5 (13.4 ± 5.3; mean ± standard deviation) points. BARS total ranged from 0–25 (10.6 ± 6.5) points for included participants, reflecting a large overlap. However, the difference in average overall disease severity was statistically significant (d=0.49, p=0.008). Excluded subjects also had higher BARS oculomotor scores (d=0.45, p=0.007). The χ^2 test for independence was used to evaluate differences in the presence of saccadic intrusions and in the presence of nystagmus between included and excluded participants. Effect sizes were assessed using φ. The presence of saccadic intrusions was only weakly associated with exclusion from the study (φ=0.26, p=0.005). The presence of nystagmus was not associated with exclusion from the study (φ=0.05, p=0.600). Of the 45 participants excluded from the study, only seven exclusions corresponded to the minimum saccade and sweep criteria. Six out of seven of these participants had ataxia and all six had a BARS oculomotor score of at least 1.0.

Computation of Horizontal Angular Gaze Velocity

Let $x(t)$ and $r_x(t)$ be the horizontal gaze position (px) and horizontal angular resolution (px/°) at time t, respectively. The derivative of the horizontal gaze position with respect to t, $x'(t)$, was numerically computed using a second-order central differences method. The angular horizontal gaze velocity (°/s) was computed as $x'(t)$ / $r_x(t)$.

Computation of Horizontal Angular Gaze Displacement

Given two horizontal gaze coordinates in the head-referenced (HREF) plane, x_1 and x_2, the horizontal displacement in degrees visual angle was computed as,

$$\left(\frac{180}{\pi}\right) arccos\left(\frac{F^2 + x_1 x_2}{(F^2+x_1^2)(F^2+x_2^2)}\right),$$

where $F = 15{,}000$ is a constant specifying the distance of the HREF plane from the eye. (This unit is independent of system setup, display distance, and display resolution.[36])

Model Training

Models were trained and evaluated using leave-one-subject-out cross-validation (LOSOCV). In each iteration of LOSOCV, one subject's data (i.e., all task recordings across longitudinal visits) were withheld as the testing set and remaining subjects' data were used as the training set. Feature standardization and hyperparameter selection (including random sampling of candidate regularization strengths) were independently performed within each iteration of LOSOCV. This procedure aims to mimic the generalization of a trained model to a previously unseen subject's data.

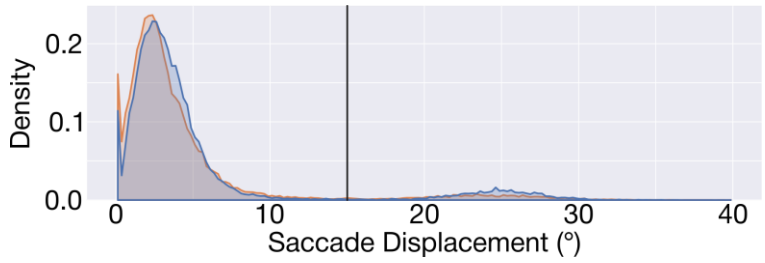
For each iteration of LOSOCV, all features were standardized such that each feature in the training set had zero mean and unit variance. Then, 100 candidate values of the regularization strength hyperparameter of Lasso were randomly sampled from the log-uniform distribution spanning 10^{-3}–10^{3}. The hyperparameter maximizing the coefficient of determination (assessed using five-fold cross validation within the training set) was used when training each model.

Ablation Study: Saccade Classes

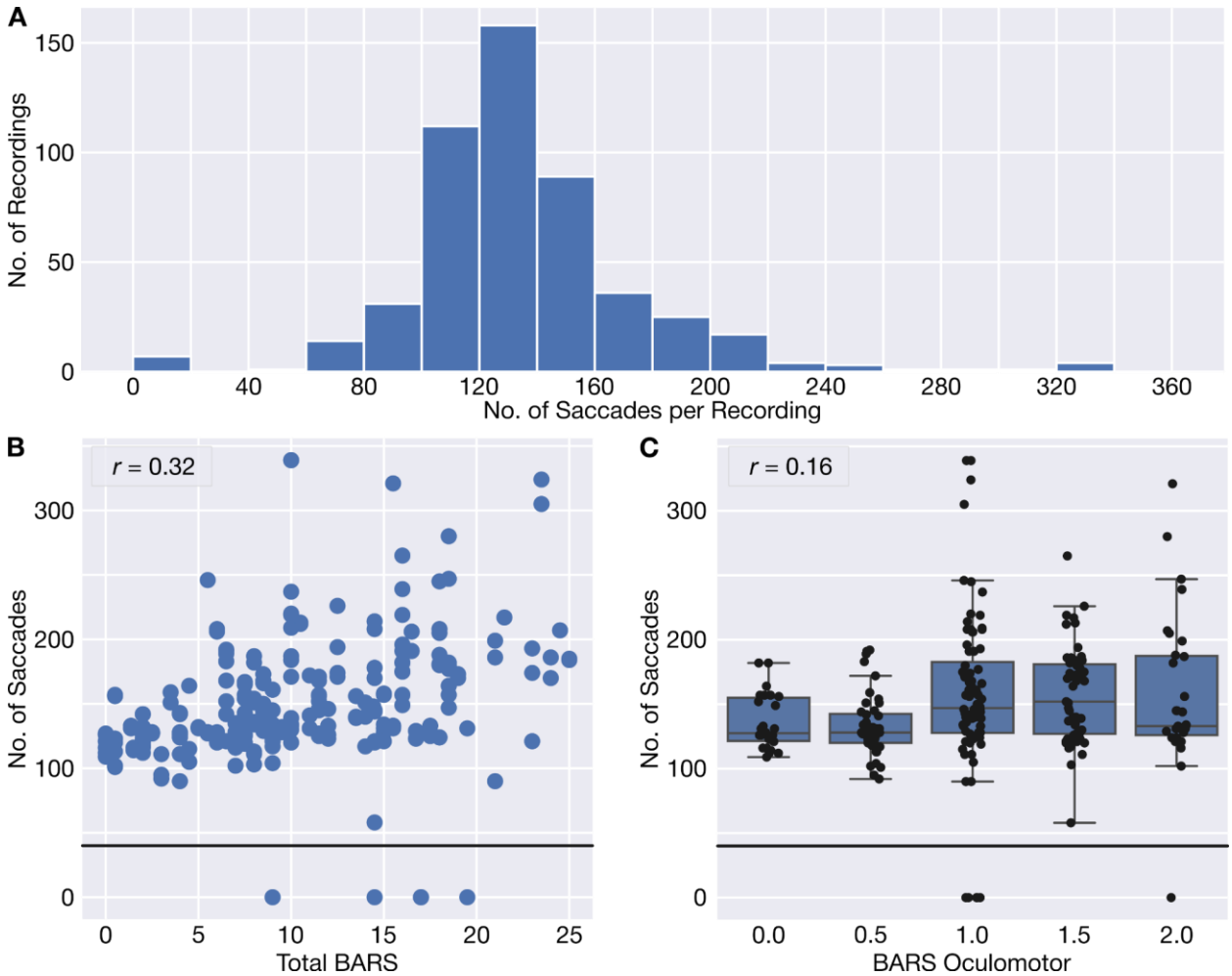
To assess the importance of extracting features for each saccade class (progressive saccades, regressions, and sweeps), we computed a set of 13 features from the set of fixations and the set of all saccades. These features were computed identically to those described in the methods, with the exception of saccades being treated as a single set and the removal of the three features capturing the percentage of each saccade class. The subsequent modeling and composite score assessment methodology remained unchanged.

The composite score produced by the ablated model demonstrated degraded performance with respect to all evaluation statistics. Though still strong, the ablated composite scores had a weaker correlation with BARS ($r=0.77$, $p<0.001$) and reduced reliability (ICC=0.93, $p<0.001$). The correlation with BARS oculomotor was slightly weaker ($r=0.50$, $p<0.001$) and the correlation with BARS speech was notably weaker ($r=0.66$, $p<0.001$). Though the ablated composite scores demonstrated only a slightly diminished capacity to distinguish between controls and participants with ataxia (AUC=0.86, $p<0.001$), the ability to detect subclinical oculomotor signs (AUC=0.64, $p=0.09$) and responsiveness to disease progression ($d=0.28$, $p=0.07$) were both adversely impacted and no longer significant.

Supplemental Figures

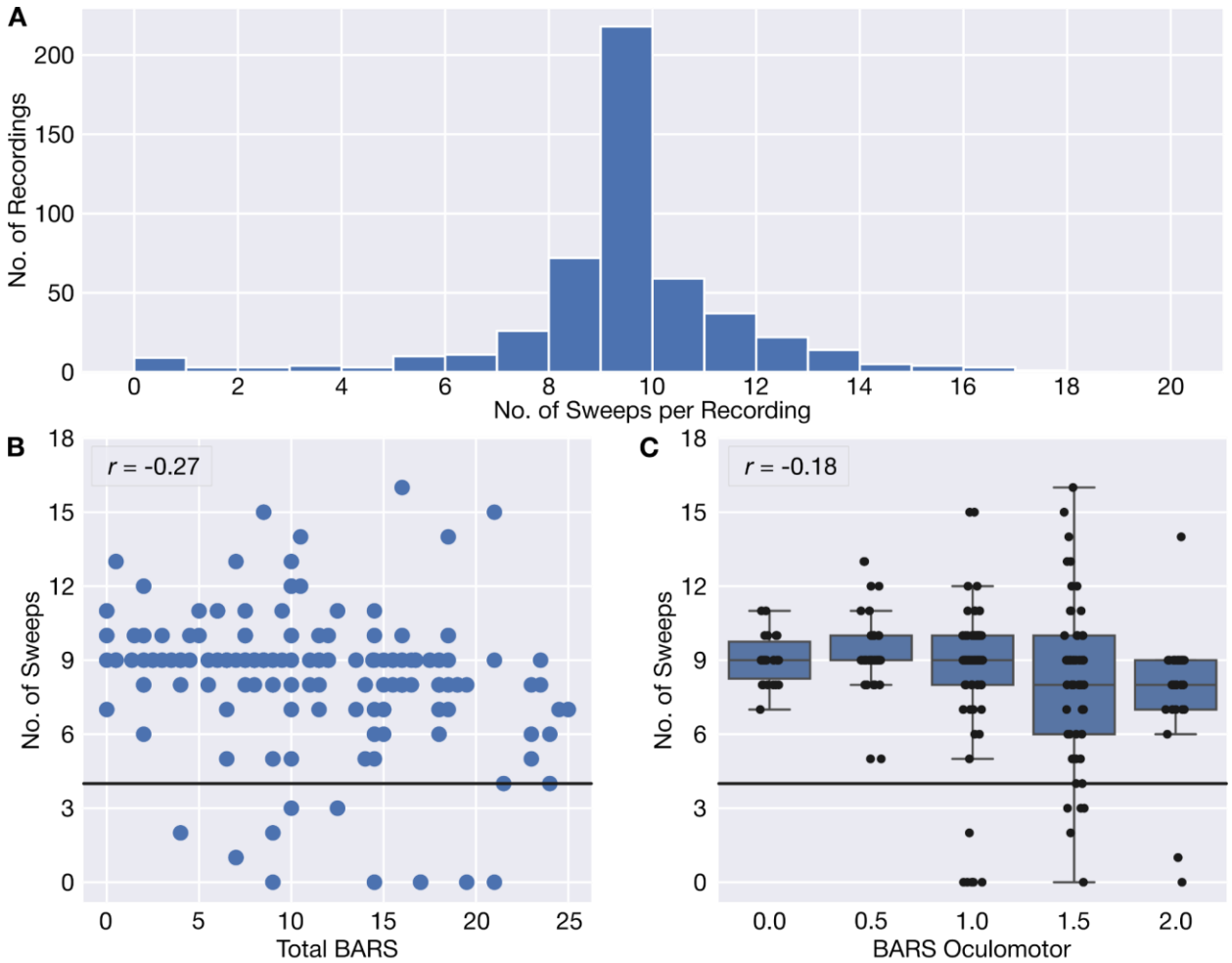


Supplementary Fig. 1. Probability density of saccade displacements. The orange curve denotes data from participants with ataxia and the blue curve denotes data from healthy controls. The black vertical line denotes the chosen 15° threshold for labeling a saccade as a sweep.

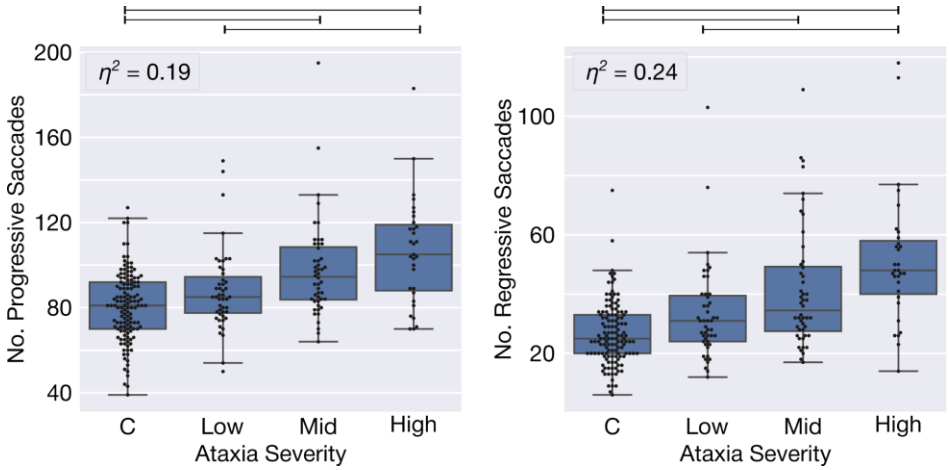


Supplementary Fig. 2. The number of saccades per task recording. (**A**) The distribution across all recordings (both participants with ataxia and healthy controls). (**B**) The relationship between clinician-assessed overall ataxia severity and the number of saccades in each task recording. The horizontal black line at 40 saccades denotes the threshold for exclusion of a recording from further analysis. (**C**) The relationship between the oculomotor component of the

Brief Ataxia Rating Scale (BARS) and the number of saccades in each task recording. Box plots denote each quartile, with a maximum whisker length of 1.5 times the interquartile range. The overlaid swarm shows each task recording.



Supplementary Fig. 3. The number of sweeps (saccades with at least 15° displacement) per task recording. (**A**) The distribution across all recordings (both participants with ataxia and healthy controls). The presented passage had ten lines of text, resulting in a peak at nine sweeps. Additional sweeps may be due to regressions to previous lines or severe undersweep. Fewer sweeps result from data quality filters (e.g., exclusion due to adjacent blinks). (**B**) The relationship between clinician-assessed overall ataxia severity and the number of sweeps in each task recording. The horizontal black line at four sweeps denotes the threshold for exclusion of a recording from further analysis. (**C**) The relationship between the oculomotor component of the Brief Ataxia Rating Scale (BARS) and the number of sweeps in each task recording. Box plots denote each quartile, with a maximum whisker length of 1.5 times the interquartile range. The overlaid swarm shows each task recording.



Supplementary Fig. 4. The number of progressive and regressive saccades per task recording, stratified by disease severity. These measures are calculated and presented in the same manner as the kinematic measures in Fig. 2.

Supplemental Tables

Supplementary Table 1. Model Features. The relative importance of each feature comprising the learned composite score was determined using the magnitude of the model coefficients. The mean and standard deviation of model coefficients reported are computed over all iterations of leave-one-subject-out cross-validation. The five most important features are bolded.

Abbreviations: C.V.—Coefficient of Variation.; Pct.—Percentage.

Subset	Feature	Importance
Fixations	**Displacement Mean**	**2.38 ± 0.04**
Fixations	Displacement C.V.	0.07 ± 0.03
Fixations	Duration Mean	0.00 ± 0.01
Fixations	**Duration C.V.**	**1.25 ± 0.07**
Fixations	Peak Speed Mean	-0.38 ± 0.03
Fixations	Peak Speed C.V.	-0.06 ± 0.06
Progressive Saccades	Pct. Progressive Saccades	0.00 ± 0.00
Progressive Saccades	Displacement Mean	-0.02 ± 0.04
Progressive Saccades	Displacement C.V.	-0.12 ± 0.07
Progressive Saccades	Duration Mean	-0.00 ± 0.02
Progressive Saccades	Duration C.V.	0.00 ± 0.00
Progressive Saccades	Peak Speed Mean	-0.05 ± 0.03
Progressive Saccades	Peak Speed C.V.	0.00 ± 0.01
Regressions	**Pct. Regressions**	**1.09 ± 0.04**
Regressions	Displacement Mean	0.44 ± 0.05
Regressions	Displacement C.V.	0.00 ± 0.01
Regressions	Duration Mean	0.00 ± 0.01
Regressions	Duration C.V.	-0.08 ± 0.05
Regressions	Peak Speed Mean	-0.04 ± 0.04
Regressions	Peak Speed C.V.	0.04 ± 0.04

Sweeps	**Pct. Sweeps**	**-2.67 ± 0.03**
Sweeps	Displacement Mean	-0.03 ± 0.02
Sweeps	Displacement C.V.	0.40 ± 0.03
Sweeps	Duration Mean	-0.70 ± 0.07
Sweeps	Duration C.V.	0.05 ± 0.04
Sweeps	Peak Speed Mean	0.00 ± 0.01
Sweeps	Peak Speed C.V.	0.41 ± 0.06
All Saccades	**Saccade Frequency**	**-1.90 ± 0.03**