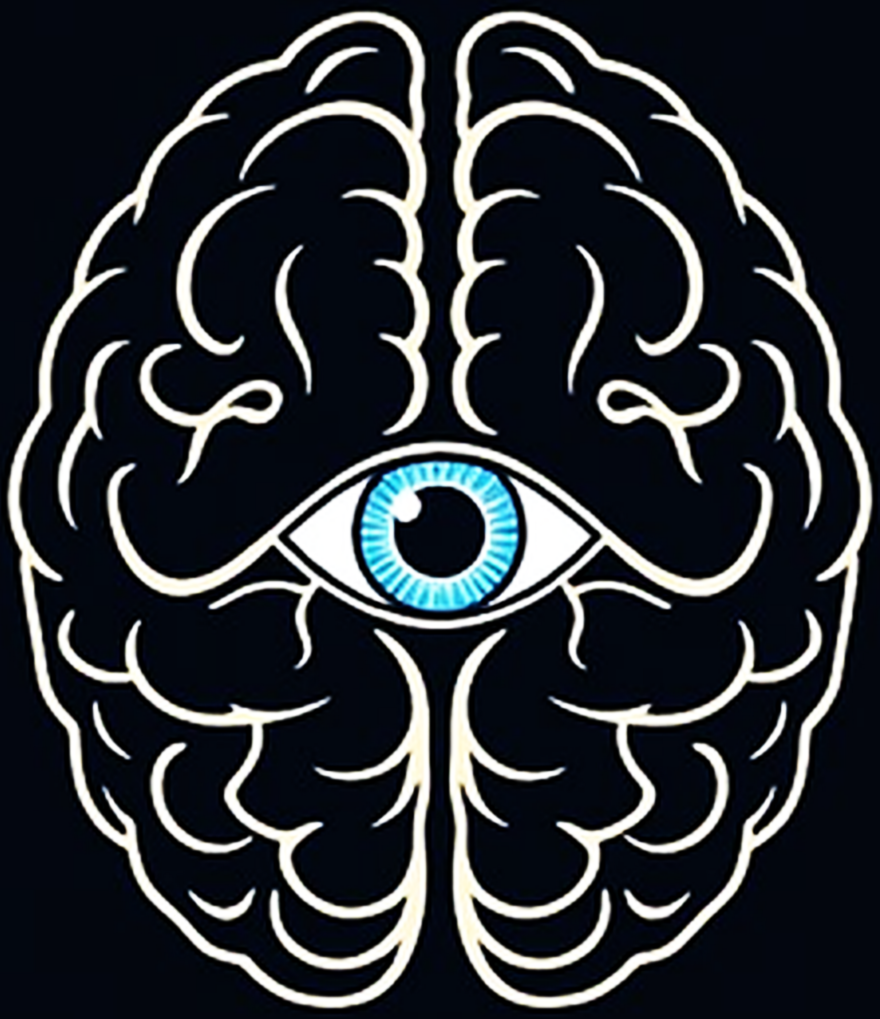
Visual Cues to Autism

How Gaze Behaviour Reflects Autistic Traits in Children

Erasmus MC, Department of Neuroscience
Ingrid Brugman



VISUAL CUES TO AUTISM

How Gaze Behaviour Reflects Autistic Traits in Children

Ingrid Brugman
Student number : 4823222

Thesis in partial fulfillment of the requirements for the joint degree of

Master of Science

in Technical Medicine

Leiden University; Delft University of Technology; Erasmus University Rotterdam

Master thesis project (TM30004 ; 35 ECTS)
Department of Neuroscience, Erasmus MC
March 3, 2025 - September 20, 2025

Supervisors:

Prof. dr. Maarten Frens	Technical Supervision
Dr. ir. Rick van der Vliet	Medical Supervision

Thesis committee members:

Prof. dr. Maarten Frens	Erasmus MC	Chair
Dr. ir. Rick van der Vliet	Erasmus MC	Committee member
Dr. Marit Ruitenberg	Leiden University	Committee member

Universiteit Leiden

TU Delft

Erasmus University Rotterdam

Preface

This thesis marks not only the completion of my masters Technical Medicine, but also my time as a student. It has been an incredible journey and I am most grateful for the people that were with me on the way. Your unconditional support and encouragement made it a way easier journey, and the fun and laughter we shared a far more enjoyable one.

I would like to thank my supervisors, Rick, Maarten, and Suzanne, for their valuable feedback and ideas that shaped this thesis into what it is today. Our Wednesday sparring sessions were both enjoyable and valuable to me. I wish you continued success and joy, both inside and outside the hospital walls.

Ingrid
September 2025

Summary

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterised by deficits in social interaction and restrictive and repetitive behaviours. ASD is often accompanied by anxiety, elevated rates of depression, and reduced quality of life. Current diagnostic tools are time-consuming, subjective, and less effective in certain groups, particularly young children. Eye tracking has emerged as a promising biomarker, holding the potential to improve diagnostic accuracy and developmental outcomes, thereby reducing the lifetime socio-economic costs of autism. This thesis aims to advance the understanding of gaze behaviour as a biomarker for ASD, thereby driving the development of more accurate, accessible, and scalable diagnostic tools.

A literature review identified fifteen eye tracking features that are associated with ASD diagnosis or Social Responsiveness Scale (SRS) scores. These features were extracted from raw eye tracking data in the Generation R database. Statistical analyses were performed using Generalised Linear Models to evaluate their relationship with ASD diagnosis and SRS scores. Additional analyses addressed feature distributions, gender differences, and the effects of video content and participant age and gender. Features with significant relationships with ASD diagnosis or SRS scores were used in a nested cross-validation framework with predictive machine learning models. The Area Under the Curve (AUC) was the primary performance metric, supported by F1 score, precision, and recall.

Classification proved challenging due to limited discriminative power of individual features. The CatBoost Gradient Boosting Decision Tree achieved the highest performance for ASD prediction with an AUC of 0.71, indicating that gaze-derived features hold promise when used in complex non-linear models. In contrast, the model predicting SRS scores performed worse, with an AUC of 0.57, suggesting that social responsiveness is a part of ASD that may be more effectively masked in adolescents. An alternative explanation is that the features reflect aspects of the autism phenotype that are unrelated to social responsiveness.

Key limitations include the small number of 33 participants with an ASD diagnosis, which reduced statistical power. The reliance on pre-existing clinical diagnoses and SRS scores, which does not capture the full complexity of ASD, led to a lack of ground truth. The adolescent age of the participants posed limitations, as masking behaviours can obscure gaze-based markers. Co-occurring factors such as attention difficulties or cognitive ability were not controlled for. Finally, choices in feature engineering, such as gaze data aggregation, reduced temporal detail; and exclusion of participants with very low screentime may have caused individuals with low social engagement to be underrepresented.

In summary, gaze analysis is a promising tool for understanding and identifying ASD. The findings in this thesis suggest that while individual gaze features offer limited diagnostic power on their own, their integration within more advanced models holds potential to improve the diagnostic process and provide deeper insights into the mechanisms underlying ASD. As datasets grow in size and quality, and stimuli continue to evolve, there are substantial opportunities to uncover more nuanced relationships between gaze behaviour and ASD traits. These developments ultimately support the creation of low-cost, inclusive and scalable ASD assessment tools, thereby enhancing both individual quality of life and broader socio-economic outcomes.

Contents

Nomenclature

ADI Autism Diagnostic Interview

$ADOS$ Autism Diagnostic Observation Schedule

AOI Area Of Interest

ASD Autism Spectrum Disorder

AUC Area Under the Receiver Operating Curve

CI Confidence Interval

DTW Dynamic Time Warping

EEG Electroencephalography

$GEE - GLM$ Generalised Estimating Equation – Generalised Linear Model

LDA Linear Discriminant Analysis

PCA Principal Component Analysis

SCL Skin Conductance Level

SRS Social Responsiveness Scale

TD Typically Developing / Typical Development

VIF Variance Inflation Factor

1

Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that affects approximately 1 in 100 children worldwide. [1] Core characteristics of ASD include deficits in social interaction and restricted, repetitive behaviours. [2] Research indicates that autistic adults experience anxiety, depression, and stress levels comparable to individuals diagnosed with primary anxiety and depressive disorders. [3] Moreover, the quality of life for individuals with ASD is consistently lower than that of the general population, across all age groups. [4, 5]

Despite the high prevalence and substantial impact of ASD across the lifespan, current diagnostic approaches face several critical limitations. Limitations of current gold-standard diagnostic tools, such as the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview (ADI), are particularly evident among certain subgroups. [6] For example, culturally and linguistically diverse populations often experience diminished diagnostic precision with these assessments. [7] Additionally, current tools frequently fail to adequately detect high-functioning autism, as they rely heavily on overt behavioural markers that may not capture the subtle social deficits of this subgroup. [8] Furthermore, these tools are not applicable in very young children, due to reliance on caregiver recall and the substantial developmental variability in children under the age of three. [9] Early identification allows for early intervention, which has been shown to significantly enhance developmental outcomes and quality of life for individuals with ASD. [10, 11] In addition to individual benefits, early intervention also yields significant economic benefits over time. [10, 12] Consequently, there is a need for objective biomarkers to complement and enhance the ASD diagnostic process.

To identify objective biomarkers for ASD, it is essential to understand its underlying neurobiological and psychological mechanisms. Although ASD is widely recognized as a neurodevelopmental condition, no single model has yet captured the full complexity of its etiology. Increasing evidence points to a multifaceted interplay of genetic, neurological, environmental, and psychological factors. [13–16] This complexity suggests that a viable biomarker must be sensitive to both the neurological and cognitive aspects of the condition.

Gaze behaviour is particularly valuable in this context, as it integrates lower-level neural processes with higher-order cognitive functions. Atypical gaze patterns have been linked to altered activity in key brain regions involved in emotion processing, attention, and social cognition, including the amygdala, frontal eye fields, dorsolateral prefrontal cortex, insula, superior temporal sulcus, and temporoparietal

junction. [17–19] At the same time, gaze behaviour is shaped by higher-order psychological processes such as motivation, joint attention, and theory of mind. This dual anchoring in both neural and cognitive domains positions gaze behaviour as a meaningful marker of ASD-related differences. Eye tracking provides a non-invasive and quantifiable method for these behaviours, and it is particularly well-suited for assessing young children. [20, 21] Gaze-related differences between individuals with ASD and those with typical development (TD) have been observed as early as infancy, which underscores the potential for earlier diagnosis. [22] Eye tracking also enables more objective assessments in the other subgroups that are underserved by traditional diagnostic tools, such as high-functioning individuals and culturally diverse populations. [23, 24] These strengths make it a promising tool to improve the diagnostic process of ASD.

To fully realise these advantages, careful consideration must be given to the type of stimulus used in eye tracking assessments. Research suggests that dynamic social stimuli involving more than one person are optimal for distinguishing individuals with ASD from TD. [25, 26] The differences between these groups become more pronounced when video-based stimuli are used, as opposed to live social interactions. [27] Moreover, naturalistic scenes yield greater ecological validity than animated characters. [28, 29] Given these findings, video recordings of naturalistic social interactions involving multiple individuals appear to be the most effective method for identifying gaze patterns associated with ASD. However, despite their promise, the specific eye-tracking outcome parameters most predictive of ASD remain unclear, as much of the existing research has focused on static stimuli.

This thesis aims to address this knowledge gap by identifying and evaluating gaze-based outcome parameters from dynamic, naturalistic social stimuli that may serve as objective biomarkers for ASD. A comprehensive literature review, detailed in Appendix A, was conducted to identify outcome parameters. These parameters will be extracted and evaluated using an eye tracking dataset, with the ultimate goal of informing more inclusive and effective diagnostic practices.

2

Methods

2.1. Data Acquisition

Participants

This study was conducted within the framework of Generation R, a longitudinal population-based cohort in Rotterdam, the Netherlands. Generation R is a longitudinal population study that follows approximately 10,000 children from the fetal stage through adulthood to study a variety of factors related to health and disease. When the children were approximately 13 years old they were invited to visit the research centre for an eye tracking task and evaluation of the Social Responsiveness Scale (SRS) short form. Between April 2016 and January 2020, a total of 3,008 children underwent the eye tracking task, all of whom had normal or corrected-to-normal vision. Valid eye tracking data was obtained from 2,920 children (97% of those tested). The majority of these children watched three videos and therefore provided three measurements (see Paragraph *Video Stimuli*), while 48 participants watched only one or two due to technical difficulties or non-participation. Individual measurements were excluded when the quality of the eye tracking data was deemed inadequate, defined as an error margin exceeding 1.5° of visual angle or a tracking percentage below 25% of the video duration. When a participant had measurements eligible for inclusion, as well as measurements meeting exclusion criteria, only the measurements meeting exclusion criteria were removed. A total of 254 observations from 148 participants was excluded because their tracking percentage fell below the 25% threshold. An additional 746 observations from 262 participants were discarded because their calibration error exceeded 1.5°. Lastly, to ensure that analyses focused on individual differences, one randomly selected twin from each twin pair who completed the eye-tracking task was excluded. This led to the exclusion of 76 children. This step was carried out after inspection of eye tracking data quality was to avoid excluding the twin that had acceptable data quality while retaining the twin with insufficient data quality, ensuring minimal data loss. The final data set consisted of 7508 measurements derived from 2539 children.

Social Responsiveness

Autistic traits were assessed using both dimensional and categorical measures: the SRS short form and clinical ASD diagnoses. The SRS short form is a validated 16-item questionnaire that assesses social functioning and traits associated with ASD, providing a total score that reflects the severity of social impairments. A lower score indicates better social functioning. This measure was administered consistently across the entire cohort. In contrast, ASD diagnoses were included only when participants

had independently sought psychological help outside the study context and received a formal diagnosis of ASD. The dual use of both measures aimed to enhance the comprehensiveness of the assessment: while the SRS offers broad and systematic coverage of autistic traits, formal diagnosis serves as a clinically grounded complement. Notably, the association between SRS scores and ASD diagnosis has been reported to be moderate ($\phi = 0.33$), suggesting that the SRS likely captures only a portion of the broader ASD phenotype. [30]

Eye Tracking

Video Stimuli

The video stimuli presented during the eye tracking task were recorded using a professional-grade digital camera. Each clip was 36.5 by 27.5 cm in size, corresponding to 40.1° x 30.8° of visual angle when viewed from a distance of 65 cm, and had a resolution of 800 x 600 pixels. Eight distinct video clips were developed, each depicting a dyadic interaction between two same-gender actors aged 20 to 25 years. The emotional content of the clips was categorised into four types: happiness, anger, sadness, and depression. All emotional contents were portrayed once by male actors and once by female actors. To ensure consistency, all scenes were recorded with a stationary camera and without cuts. Actors wore neutral clothing and stood in fixed positions against a plain backdrop. A poster was included in the background of each scene as a non-social object. The emotional scripts were carefully developed in collaboration with a local film director to ensure authenticity, realism, and emotional clarity. Each dialogue maintained a standardised structure in terms of sentence length, pauses, and turn-taking, though the emotional content varied. In each interaction, one actor expressed the emotion while the other remained neutral. Participants were exposed to a random selection of three out of eight clips, each depicting a different emotion. The actors were in at least one of the three videos female and at least one of the three videos male.

Apparatus and procedure

Eye tracking data was collected in a quiet, standardly lit room using the SMI Red250 remote eye tracking system. The device recorded gaze data at a sampling rate of 120 Hz. The system provided a spatial resolution of 0.1° and gaze accuracy of up to 0.5°, while permitting limited head movement within a defined three-dimensional space (32×21×30 cm) at a distance not exceeding 75 cm. Participants were not restrained by a chin rest to allow for naturalistic viewing behaviour. Before the experiment commenced, a 9-point calibration procedure was performed. Participants were asked to fixate on a black dot that appeared sequentially at nine locations on the screen. Calibration was considered successful when the average error remained below 1.5° of visual angle.

2.2. Data Processing

2.2.1. Speech Detection

Speech detection was conducted using Python version 3.13.2, which enabled frame-wise energy computation of audio signals of the librosa library. Speech segments were identified based on manually adjustable energy thresholds that were uniquely defined for each video to accommodate differences in recording conditions and speaker characteristics. The initially detected segments were further processed by merging adjacent segments separated by small gaps to accommodate short silences between words, using video-specific max_gap values. Subsequently, the frame numbers were manually reviewed against the video clips to confirm the final onset and offset frame numbers. The resulting speech segment boundaries, expressed in frame indices, were converted to time (in seconds) by dividing the frame numbers by the video frame rate of 25 fps.

2.2.2. Feature Extraction

Feature extraction was conducted using custom-written MATLAB scripts, using MATLAB version 2024B. Note on all metrics: If only one eye was successfully tracked, the feature values were based solely on the data of the tracked eye. Features were chosen guided by the literature review conducted prior to the start of this thesis, which is attached in Appendix A. Adjustments included the removal of the ratio between speaker and non-speaker fixation, which was justified by the major mathematical issues related to the use of ratios. [31] In order to still include information about fixation on the non-speaker during speech, the gaze-speech time and silhouette correlation were computed for the non-speaker fixation, in addition to the speaker fixation.

Screen Time

To estimate the screen engagement of a participant, a screen time value was computed. The parameter utilised for this purpose consisted of a number ranging from zero to one, denoting the percentage of the video duration during which the eye movements of a participant occurred within the screen that was displaying the video. This value was available for both the left and the right eye. In instances where both eyes were successfully tracked, the screen time value was calculated as the mean of the two:

$$\text{ScreenTime} = \frac{1}{2}(Q_{\text{left}} + Q_{\text{right}})$$

Gaze-Speech Time Correlation

To assess the temporal alignment between speech and fixation on the speaker and non-speaker during speech, the Pearson correlation coefficient was computed. The signals were represented as binary square waves. The speech square wave encoded the presence (1) or absence (0) of speech in the video. The other square wave indicated fixation on either the speaker or non-speaker during speech and was encoded as the presence (1) or absence (0) of fixations on the speaker or non-speaker during speech. The time correlation was computed twice: Once for fixation on the speaker during speech and once for fixation on the non-speaker during speech. If both eyes were successfully tracked, a frame was not registered as a fixation on the speaker or non-speaker if this was true for only one eye, as this indicates either a border or a measurement error. During silence, both square waves had value 0. An example plot of the two binary square waves is provided in Figure 2.1.

Given the two binary sequences, X and Y, the Pearson correlation coefficient r was calculated as:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2}\sqrt{\sum(Y_i - \bar{Y})^2}}$$

where $\bar{X}$ and $\bar{Y}$ represent the mean values of sequences X and Y, respectively.

The numerator of the Pearson correlation coefficient is given by:

$$\sum(X_i - \bar{X})(Y_i - \bar{Y})$$

which represents the covariance between X and Y, quantifying their joint variation.

The denominator consists of the standard deviations of X and Y, given by:

$$\sqrt{\sum(X_i - \bar{X})^2} \quad \text{and} \quad \sqrt{\sum(Y_i - \bar{Y})^2}$$

respectively, and they measure the extent of variation within each individual sequence.

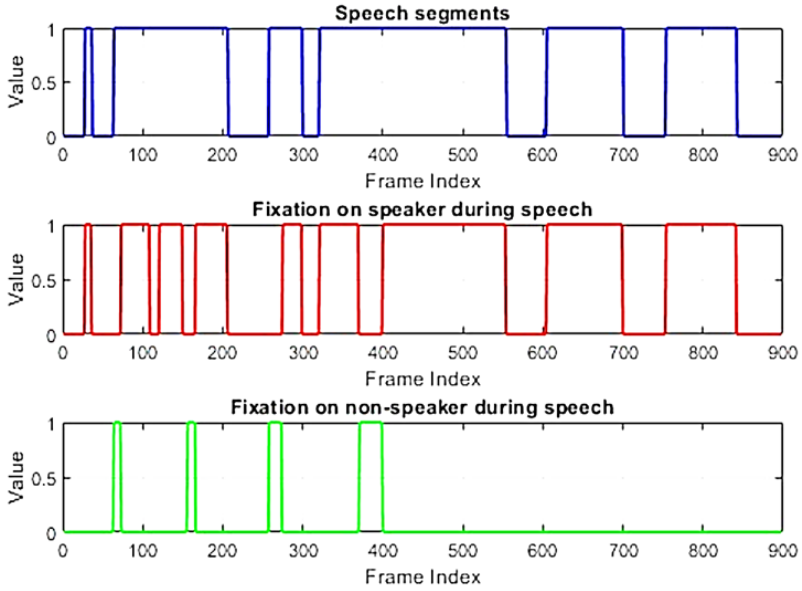


Figure 2.1: Example of the binary square waves used for the calculation of gaze-speech time and silhouette correlation. The first square wave indicates speech segments. The second square wave indicates fixation on the speaker during speech. The third square wave indicates fixation on the non-speaker during speech.

By combining the covariance and standard deviations, the correlation coefficient r quantifies how strongly deviations in X correspond to deviations in Y. Theoretically, Pearson's correlation coefficient r may range from -1 to 1. However, as the fixation square wave only contains a value of one in situations where the speech square wave is also one, the value of r can only range from 0 to 1 in this application.

If the fixation square wave was identical to the speech square wave, the correlation coefficient was set to 1, indicating perfect alignment. Conversely, if the fixation square wave contained only zeros, the coefficient was set to 0, reflecting a total lack of association.

Gaze-Speech Silhouette Correlation

The silhouette correlation between the same square waves provides an additional perspective on the alignment between the two sequences. Again, the silhouette correlation was computed twice: Once for fixation on the speaker during speech and once for fixation on the non-speaker during speech. Unlike time correlation, which focuses on the temporal synchronisation of events, the silhouette correlation emphasises their shape and structural similarity.

In order to illustrate the distinctiveness of time and silhouette correlation, consider two individuals who engage with the speaker in different ways during speech. In the first case, the participant maintains an uninterrupted gaze on the speaker for half the duration of the speech. In the second case, the participant takes multiple short glances, each of which, when summed, also accounts for half of the speech duration. In both instances, time correlation would yield the same value. However, the silhouette correlation provides an alternative perspective, as it takes into account the structure of gaze behaviour. The short glances of the second participant would result in a lower silhouette correlation due to the way the Dynamic Time Warping (DTW) algorithm accounts for shifts, stretches, and compressions in the gaze sequence. The silhouette correlation is penalised more heavily in the case of short glances, due to a higher $D_{\min}$ value, as the DTW algorithm involves a cumulative cost matrix that increases when more

adjustments are required to align the sequences. Therefore, time and silhouette correlation capture complementary aspects of gaze behaviour.

The silhouette correlation was computed using DTW with a shifting window. The frame rate of the videos was 25 fps. Previous research has shown that approximately 200 ms is required to initiate a saccade. [32] Based on this, the maximum allowed shift was set to five frames, as this shift accommodates the time needed for a saccade to either initiate towards or away from the speaker or non-speaker.

The minimum DTW distance over all shifts was given by:

$$D_{\min} = \min_{s \in [-m,m]} DTW(X_s, Y_s)$$

where X_s and Y_s represent the shifted sequences. The silhouette correlation C_s was then defined as:

$$C_s = 1 - \frac{D_{\min}}{\max(N_X, N_Y)}$$

where N_X and N_Y denote the sequence lengths.

The silhouette correlation C_s may range from 0 to 1, where 1 indicates perfect similarity, and 0 indicates no similarity. The formula is typically normalised by dividing the DTW distance by the maximum sequence length to scale the result. However, since the sequences in this analysis are of equal length, the silhouette correlation can be expressed as:

$$C_s = 1 - D_{\min}$$

Scanpath Variance

This metric provides a measure of how much the participant's fixation points deviate from the average fixation point of all participants throughout the clip. A higher average Euclidean distance between a participants fixation point and the average fixation point of all participants indicates a more atypical scanpath. The Euclidean distance was calculated for each frame where a participant exhibited a fixation. Subsequently, these Eucleadian distances were averaged. The average Euclidean distance D_{avg} can be computed using the formula:

$$D_{avg} = \frac{1}{N} \sum_{i=1}^{N} \sqrt{(X_i - X_{avg})^2 + (Y_i - Y_{avg})^2}$$

where:

- N is the number of fixations,
- X_i, Y_i represent the X- and Y-coordinates of the individual fixation points,
- X_{avg}, Y_{avg} represent the average X- and Y-coordinates for the corresponding frame.

Predictive Saccades

Predictive saccades reflect the ability to anticipate social cues. A predictive saccade was defined as a saccade towards the next speaker during the silence period between speech turns. If both eyes were successfully tracked, a saccade was not classified as towards the next speaker unless this was true for both eyes. The time window was shifted by 200 ms, or five frames, as this is the duration

required to initiate a saccade. [32] Given that each video contained six turns of speech, there were five silence periods in which a participant could exhibit a predictive saccade. Consequently, the number of predictive saccades varied between 0 and 5.

Time per Fixation

The average time per fixation was computed in frame numbers using the following formula:

$$\text{avgFixationDuration} = \frac{\sum_{i=1}^{n} \text{Duration}_i}{\text{n}}$$

Saccadic Speed

The saccadic speed was registered by the eye tracking apparatus. To correct for the relationship between amplitude and velocity in saccades, the saccadic speed was divided by the amplitude to obtain the slope. [33] The slope which was converted to the a measure for the average saccadic speed using the following formula:

$$\text{avgSaccadicSpeed} = \frac{\sum_{i=1}^{n} \frac{\text{Speed}_i}{\text{Amplitude}_i}}{\text{n}}$$

Fixation Durations on Areas of Interest

To quantify attention towards specific AOIs in the video stimuli, the total fixation duration per object category was computed as the percentage of the analysed time in that video that was spent fixating on a certain AOI. The AOIs included were body, eyes, face, mouth, object and background. To accommodate calibration inaccuracies and borderline fixations, a buffer zone of 1.5° of visual angle was applied to each area of interest (AOI), following the suggestions of previous work in the field. [34–36] The face AOI consisted of the eyes, mouth and head. The background AOI consisted of the background within the video and the black bars on the screen. The body AOI consisted of the body and the neck. If both eyes were successfully tracked, a fixation was not considered to be directed at a certain AOI if this was true for only one eye, as this indicates either a border or a measurement error.

2.2.3. Feature Inspection

Feature Distribution

To comprehensively understand the properties of the measured features, three distribution analyses were conducted: (1) Population-wide distribution fitting, (2) Diagnosis-specific distribution fitting, and (3) Feature association with SRS scores. Firstly, population-level distributions were assessed by fitting a set of candidate parametric probability density functions (PDFs) to each feature across the full cohort. The distributions considered included the normal, gamma, beta, and uniform distribution families. The gamma and beta distributions were only considered when the feature values fell within their respective valid ranges: [0, ∞] for the gamma distribution and [0, 1] for the beta distribution. Parameters for each candidate distribution were estimated using maximum likelihood estimation, and the goodness-of-fit was assessed by the log-likelihood function, defined as:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \log f(x_i \mid \theta),$$

where $f(x_i \mid \theta)$ denotes the PDF evaluated at observation x_i given parameters θ. The distribution yielding the maximum log-likelihood was selected as the best-fitting model. The results were visualised through kernel density histograms overlaid with the fitted PDFs.

Secondly, feature distributions were analysed stratified by autism diagnosis (0 = non-autistic, 1 = autistic). Within each diagnostic subgroup, the same distribution was fitted. The resulting distributions were visualised both in separate panels and via overlaid plots, enabling direct visual comparison of how the feature distributions differ between diagnostic groups.

Finally, the relationship between features and the continuous SRS scores was examined. Scatterplots were constructed with feature values plotted against SRS scores, coloured by diagnosis group. Separate linear regression models were fitted within each diagnostic group to assess differences in the relationship between features and social responsiveness.

Together, these analyses provided a foundation for understanding feature behaviour both at the population level and across clinically relevant subgroups, leveraging parametric modelling and visualisation techniques.

Control of Video, Gender, and Age Effects on Features

To examine the influence of stimulus content and participant characteristics on the features, each feature was analysed to see whether its values changed depending on the content of the video being watched, participant gender, and participant age. For video- and gender-related effects, mean and 95% confidence intervals were calculated per clip and visualised to assess variability, followed by one-way ANOVAs and Benjamini-Hochberg-corrected p-values. Effect sizes were quantified using η², and significant effects were corrected by centring feature values within clips or genders. Furthermore, gender-related effects on SRS scores were examined, as previous research has demonstrated a significant association between gender and raw SRS scores [37]. One participant with missing gender data was retained without gender-specific centring. Age-related effects were examined using linear regression with age entered as a continuous predictor. Scatterplots with fitted regression lines were generated to visualise the associations, and statistical significance was evaluated based on regression coefficients and their Benjamini-Hochberg corrected p-values. Significant effects were removed by replacing original values with residuals from the regression model, effectively eliminating linear age-related variance. Pseudo eta squared was used to quantify effect sizes using the formula:

$$\eta^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

where:

- y_i is the observed value of the feature,
- $\hat{y}_i$ is the value of the feature predicted by the regression model based on age,
- $\bar{y}$ is the mean of the observed values of the feature across all participants.

Additionally, age-related effects on SRS scores were examined, as previous research has demonstrated an association between age and raw SRS scores [37]. Data from 240 participants with missing age information was retained without correction to preserve dataset integrity.

2.3. Statistical Analysis

The aim of statistical analysis was to gain insight in the relationship between the features and both autism diagnosis and SRS scores, as well as to identify candidate features for predictive models. For both autism diagnoses and SRS scores a Generalised Estimating Equations Generalised Linear Model (GEE-GLM) was employed, enabling robust modelling of the outcome while accounting for repeated

measurements within participants. The linearity assumption of the model was verified by visually inspecting the residuals, which were centered around zero. Due to the presence of extreme outliers in both the speaker and non-speaker silhouette correlation features, these variables were winsorised at the 1st and 99th percentiles. Features and SRS scores were scaled using z-score scaling for interpretability of regression coefficients. For the autism diagnosis outcome the family was set to binominal with a logistic link function. To choose the best-fitting family for the SRS scores, both gaussian and gamma families were tested, and the one with the lowest average QIC score was selected. Consequently, the Gaussian family with an identity link function was used. To achieve better model fit, the SRS scores were transformed using a rank-based inverse normal transformation. Participant ID was specified as the clustering variable to model within-subject correlations. Wald tests from the GEE-GLM models were used to evaluate statistical significance, with p-values adjusted using the Benjamini-Hochberg procedure to control for multiple testing. A pseudo eta squared effect size was calculated using the following formula:

$$\eta^2 = \frac{D_{\text{null}} - D_{\text{full}}}{D_{\text{null}}}$$

where:

- D_{null} is the deviance of the intercept-only model, representing model fit without the predictor,
- D_{full} is the deviance of the full model including the predictor.

In addition to the ASD model including all participants, a separate ASD model was estimated using only male participants. This decision was based on the well-documented underdiagnosis of ASD in females, a pattern that is also evident in the present dataset: only 4 out of 33 participants diagnosed with ASD are female. Given this substantial imbalance, restricting the analysis to males enhances the reliability of the model estimates. [38]

Furthermore, the interacting factor gender was tested for each feature. This approach was motivated by the well-documented differences in how ASD symptoms manifest across genders. [39–41] Consequently, it is plausible that distinct models are needed to accurately predict social responsiveness and ASD for each gender, potentially even including different subsets of features. Due to the limited number of female participants with ASD, 4 out of 33, conducting gender-specific analyses for diagnosis was not feasible, underscoring the likelihood of underdiagnosis in our cohort. Therefore, the gender comparison will be restricted to models related to SRS scores.

2.4. Predictive Modelling

Feature Redundancy

Understanding and eliminating feature redundancy is essential to ensure that predictions are based on distinct, non-overlapping information, thereby reducing overfitting and enhancing model interpretability. Redundant features decrease statistical power without contributing additional information. Therefore, redundant features were removed from the dataset.

To assess feature redundancy, a correlation matrix was computed for the candidate feature set. Features were considered redundant if their correlation coefficient exceeded 0.80. The correlation between pairs of features was calculated using the Pearson correlation coefficient, defined as:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2}\sqrt{\sum(Y_i - \bar{Y})^2}}$$

where $\bar{X}$ and $\bar{Y}$ represent the mean values of sequences X and Y, respectively. This formula was previously used and explained in more detail in Subsection 2.2.2, Paragraph Gaze-Speech Time Correlation.

To gain further insight into the data, the Variance Inflation Factor (VIF) was calculated for each feature. The VIF serves as an indicator of how much the variance of a feature is inflated due to its correlation with a combination of other features in the dataset. A high VIF value suggests that a feature is highly correlated with one or more other features, and is therefore redundant. A VIF greater than 10 was considered to indicate serious multicollinearity requiring feature removal. [42] The VIF for a given feature was computed using the following formula:

$$\text{VIF}(X_i) = \frac{1}{1 - R_i^2} \tag{2.1}$$

where R_i^2 is the coefficient of determination when feature X_i is regressed on all other features. The R_i^2 value is a statistical measure that represents the proportion of the variance in the feature that is predictable from the other features. A higher R_i^2 value indicates a higher correlation between the feature and the other feature, and leads to an elevated VIF. The formula to calculate R^2 is:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where:

- y_i is the actual observed value for the i-th data point,
- $\hat{y}_i$ is the predicted value for the i-th data point based on the regression model,
- $\bar{y}$ is the mean of the actual observed values,
- n is the number of data points.

The denominator corresponds to the Residual Sum of Squares (RSS), which measures the variance of the residuals, and is expressed as:

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

The numerator corresponds to the Total Sum of Squares (TSS), which measures the total variance in the data, and is given by:

$$TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

Model Selection

Predictive models were selected based on their diversity in complexity, feature importance interpretability, and effectiveness in ranking features, all of which are key considerations for exploratory clinical research. Insight into feature importance is paramount in research contexts where understanding the underlying drivers of predictions is as critical as predictive performance itself. To meet these requirements, a diverse set of three model types was chosen, spanning a spectrum of complexity:

- Low complexity: Linear Regression (with L1, L2, and elastic net regularisation),
- Medium complexity: Random Forest,
- High complexity: CatBoost Gradient Boosting Decision Tree (CatBoost GBDT).

Data Preprocessing

To mitigate class imbalance, the dataset was downsampled to ensure an equal number of diagnosed and non-diagnosed individuals. Features were standardised using z-score normalisation for the Logistic Regression classifier. The Random Forest and CatBoost GBDT do not require z-score normalisation, as these tree-based models are scale-invariant and can handle raw feature distributions without impacting performance.

Dimensionality Reducting and Ensembling

Alternative approaches were explored with the aim of improving predictive performance. These approaches were evaluated independently, in parallel to the primary modelling strategy. The alternative strategies included:

- **Principal Component Analysis (PCA)** was applied as an unsupervised dimensionality reduction technique. It transforms the original features into a smaller set of new variables, called principal components, which capture the majority of the variance in the data.
- **Linear Discriminant Analysis (LDA)** was used as a supervised dimensionality reduction approach. Unlike PCA, LDA takes class labels into account and projects the data onto a new axis that maximises the separation between classes.
- **Ensembling** was used to combine the predictive probabilities of the three models using the raw features (Logistic Regression, Random Forest, and CatBoost GBDT) by averaging their outputs. This approach leverages the complementary strengths of each model type, improving robustness compared to any single model alone.

Nested Cross-Validation

A nested cross-validation framework was employed to obtain realistic estimates of model performance and to prevent overfitting during hyperparameter selection. The outer loop consisted of $K = 3$ stratified group folds, ensuring that all samples from a given participant were assigned to the same fold. Within each training fold, an inner $K = 3$ stratified group cross-validation was used for hyperparameter tuning, using the grid search displayed in Table 2.1. The optimal hyperparameters were used to evaluate performance on the test set of the outer loop. A visual representation of the nested cross-validation framework is provided in Figure 2.2.

Participant-level Prediction and Aggregation

Model predictions were computed at the sample level and aggregated per participant by averaging predicted probabilities:

$$\hat{p}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \hat{p}_{ij}$$

where $\hat{p}_{ij}$ is the predicted probability for the j-th sample of participant i, and n_i is the number of samples for participant i. Final binary predictions were obtained by thresholding at 0.5:

$$\hat{y}_i = \begin{cases} 1 & \text{if } \hat{p}_i \geq 0.5 \\ 0 & \text{otherwise} \end{cases}$$

Table 2.1: Hyperparameter grids used for model tuning.

Model	Hyperparameter	Values
Logistic Regression	penalty	`l1`, `l2`, `elasticnet`
	C	0.001, 0.01, 0.1, 1, 10
	l1_ratio (only for `elasticnet`)	0.2, 0.5, 0.8
Random Forest	n_estimators	50, 100, 150
	max_features	`sqrt`, `log2`, None
	max_depth	5, 10, 20, None
CatBoost	iterations	100, 200, 300
	learning_rate	0.01, 0.05, 0.1
	depth	4, 6, 8
	l2_leaf_reg	1, 3, 5
Logistic Regression (PCA)	penalty	`l1`, `l2`, `elasticnet`
	C	0.001, 0.01, 0.1, 1, 10
	l1_ratio (only for `elasticnet`)	0.2, 0.5, 0.8
	n_components	2, 4, 6
Logistic Regression (LDA)	penalty	`l1`, `l2`, `elasticnet`
	C	0.001, 0.01, 0.1, 1, 10
	l1_ratio (only for `elasticnet`)	0.2, 0.5, 0.8
Ensemble	NA	

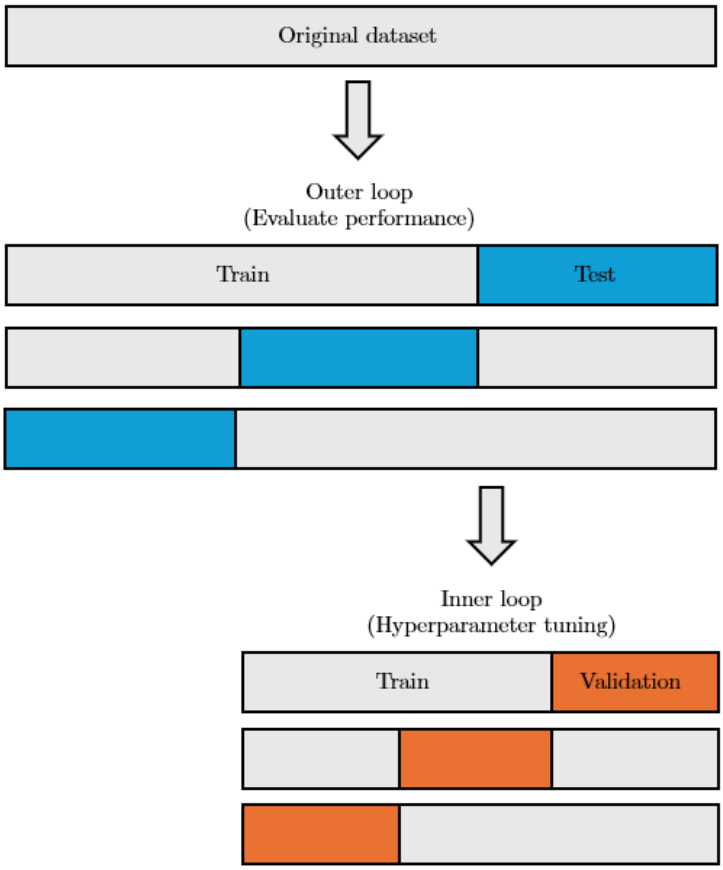


Figure 2.2: Nested cross-validation framework as applied in this study.

Performance Metrics

To evaluate model performance, the primary metric was the area under the receiver operating characteristic curve (AUC). The AUC was chosen because this work represents an early stage in the use of technical tools for ASD assessment. At this stage, the focus is on the models overall ability to discriminate between classes, rather than on optimising the trade-off between false positives and false negatives. As secondary performance metrics, the F1 score, precision and recall were computed. The F1 score is the harmonic mean of precision and recall, making it a robust measure for evaluating binary classification performance. Model performance was evaluated using these metrics, and the best performing model was chosen for predicting autism diagnosis and SRS scores separately. When two models performed similarly, the less complex model was selected for reasons of interpretability and deployment feasibility.

Definition of Reference Labels

A formal autism diagnosis was used as the reference label, distinguishing between individuals with and without ASD. In addition, SRS raw scores were used as an alternative outcome reflecting levels of autistic traits. A cut-off score of 15 was applied to classify individuals as having elevated versus typical levels of social impairment, based on findings by Nguyen et al., who identified this threshold as optimising both sensitivity and specificity. [43] However, important to note is that subsequent research by Lyall et al. has raised concerns about the validity of this cut-off in females, suggesting potential gender-related differences in SRS score interpretation. [44] Currently, no more well-established threshold than 15 for both genders exists. Therefore, this cut-off was adopted for the present analysis.

3

Results

3.1. Participant Characteristics

Table 3.1 presents descriptive statistics of the participants included in the final dataset, as described in Section 2.1. It summarises the distributions of gender, age, autism diagnosis, and SRS scores. Participants with missing information on autism diagnosis or SRS scores were excluded from analyses involving that specific outcome, but were retained for analyses of the other outcome.

Table 3.1: Participant Characteristics Summary.

Total	
Total	N = 2539
Gender	
Male (%)	N = 1218 (47.97%)
Female (%)	N = 1320 (51.99%)
Unknown (%)	N = 1 (0.04%)
Age	
μ (σ)	μ = 13.52 (σ = 0.36)
Min, Max	12.60, 16.63
Unknown (%)	N = 240 (9.45%)
Autism Diagnosis	
No (%)	N = 2502 (98.54%)
Yes (%)	N = 33 (1.30%)
Unknown (%)	N = 4 (0.16%)
SRS Score	
μ (σ)	μ = 4.76 (σ = 3.66)
Min, Max	0.0, 46.0
Unknown (%)	N = 278 (10.95%)

3.2. Data Processing

Feature Distribution

The fitted distributions are visualised in Appendix B, which provides a per-feature summary of the empirical distributions. The fitting of the full cohort distribution revealed that no single parametric family consistently provided the best fit in all features. Although the normal distribution frequently offered a reasonable approximation of central tendency, features exhibiting skewness or bounded support were more accurately modelled by the gamma or beta distributions. Log-likelihood values substantiated these differences, with the beta and gamma families outperforming the normal distribution particularly for skewed features.

When distributions were stratified by autism diagnosis, group-level similarities and differences became apparent. In several features, the ASD and TD group had a nearly fully overlapping distribution. In certain features, however, the ASD and TD group diverged in the shape and dispersion of their distributions. This phenomenon became most apparent in the face fixation feature. This feature was best modelled by a beta distribution in both groups, but exhibited greater skewness in the autistic subgroup.

It was noticed that associations between features and SRS scores visually appeared to be diagnosis-dependent. This hypothesis was confirmed with a linear regression model including interaction terms for autism diagnosis group, of which the results are displayed in Table 3.2. Eleven out of fifteen features indeed demonstrated a notably different, usually stronger, association with SRS scores within the ASD group compared to the TD group. This pattern suggests that social responsiveness may be more reliably predicted in individuals with ASD than TD. The direction of the relationship was generally aligned with expectations based on the literature review conducted prior to this thesis, which can be found in Appendix A. Interestingly, two features, object fixation and non-speaker time correlation, exhibited statistically significant interaction effects in the opposite direction for both groups. Higher values of these features were associated with lower social functioning in the TD group, whereas in the ASD group higher values were associated with increased social functioning. In the TD group, increased fixation on distracting objects may indicate reduced social engagement, substituting for socially oriented gaze like face fixation. In contrast, for individuals with autism, more object fixation may reflect partial social engagement rather than avoidance, and therefore may be a positive marker of social functioning.

Control of Video, Gender, and Age Effects on Features

The plots of the mean values of features per videoclip are provided in Appendix C. These revealed systematic variability in features depending on the video content, which was confirmed by the results of the one-way ANOVA provided in Table 3.3. Each feature was affected significantly by video content, with effect sizes ranging from 0-21%. Therefore, all features were corrected, such that each feature value was expressed relative to the mean feature value for that video. Given that the initial effect size of the video clip on the scanpath variance was observed to be excessively high ($\eta^2 = 0.21$), further investigation was conducted to understand the cause of this effect. The key contributing factor turned out to be the varying physical distance between the main characters in each video. When the spatial distance between the main characters was larger, an identical shift in fixation from one character to the other resulted in a larger Euclidean distance. To validate this confounding variable, a correction was applied to the scanpath variance measure. The ear-to-ear distance between the main characters was estimated for each video and used as a predictor in a linear regression model with scanpath variance as the dependent variable. Using the residuals from this model reduced the effect size of the video clip on scanpath variance significantly ($\eta^2 = 0.02$), suggesting that much of the original effect was indeed attributable to differences in spatial layout.

Table 3.2: Interaction p-values for diagnosis group in the relationship between each feature and SRS scores, with regression coefficients for TD and ASD groups.

Feature	p	**Coefficient TD**	**Coefficient ASD**
ScreenTime	0.00	-0.99	-37.96
SaccadicSpeed	0.00	-0.00	-0.27
FaceFixation	0.00	-1.15	-13.21
SpeakerTimeCorrelation	0.00	-1.90	-23.22
EyesFixation	0.00	-0.70	-15.63
TimePerFixation	0.00	-0.00	-0.00
PredictiveSaccades	0.00	-0.12	-1.34
ObjectFixation	0.01	1.77	-18.21
NonSpeakerTimeCorrelation	0.01	0.71	-13.11
ScanpathVariance	0.01	0.00	0.02
BodyFixation	0.01	1.08	6.62
NonSpeakerSilhouetteCorrelation	0.18	-0.87	5.87
BackgroundFixation	0.39	4.21	-0.91
SpeakerSilhouetteCorrelation	0.51	-3.47	5.64
MouthFixation	0.68	-0.48	0.10

Table 3.3: Summary of one-way ANOVAs testing the effect of video content on each feature, ordered by increasing p-value.

Feature	p	η^2
ScanpathVariance	< 0.001	0.21
SpeakerTimeCorrelation	< 0.001	0.12
PredictiveSaccades	< 0.001	0.11
SpeakerSilhouetteCorrelation	< 0.001	0.09
NonSpeakerTimeCorrelation	< 0.001	0.08
MouthFixation	< 0.001	0.07
NonSpeakerSilhouetteCorrelation	< 0.001	0.07
FaceFixation	< 0.001	0.06
TimePerFixation	< 0.001	0.03
SaccadicSpeed	< 0.001	0.03
BodyFixation	< 0.001	0.02
EyesFixation	< 0.001	0.01
ObjectFixation	< 0.001	0.01
BackgroundFixation	< 0.001	0.01
ScreenTime	< 0.001	0.00

The plots of the mean values of features and SRS scores per gender are provided in Appendix D. These revealed systematic variability in the majority of the features depending on the gender of the participant, which was confirmed by the results of the one-way ANOVA displayed in Table 3.4. All features except eye fixation were affected significantly by participant gender, with effect sizes ranging from 0-2%. All affected features were corrected for gender-related effects, such that each feature value was expressed relative to the mean for that feature in the gender group of the participant. SRS scores were corrected using the same method.

Table 3.4: Results of one-way ANOVAs testing the effect of participant gender on each feature, ordered by increasing p-value.

Feature	p	η^2
SpeakerTimeCorrelation	p < 0.001	0.02
ScanpathVariance	p < 0.001	0.02
FaceFixation	p < 0.001	0.01
SRS	p < 0.001	0.01
ScreenTime	p < 0.001	0.01
NonSpeakerTimeCorrelation	p < 0.001	0.01
MouthFixation	p < 0.001	0.01
SaccadicSpeed	p < 0.001	0.00
BackgroundFixation	p < 0.001	0.00
ObjectFixation	p < 0.001	0.00
PredictiveSaccades	p < 0.001	0.00
TimePerFixation	p < 0.001	0.00
BodyFixation	p = 0.01	0.00
NonSpeakerSilhouetteCorrelation	p = 0.02	0.00
SpeakerSilhouetteCorrelation	p = 0.03	0.00
EyesFixation	p = 0.09	0.00

Scatterplots showing the linear regression of features and SRS scores across the continuous participant age range are provided in Appendix E. The summary of the regression analyses is shown in Table 3.5. No features or SRS scores demonstrated statistically significant relationships with age. Therefore, no age-related corrections were applied.

Table 3.5: Regression results for the effect of participant age on each feature, ordered by increasing p-values.

Feature	**Coefficient**	p	η^2
SpeakerSilhouetteCorrelation	0.003	0.11	0.00
MouthFixation	-0.020	0.11	0.00
BackgroundFixation	0.004	0.12	0.00
SRS	0.268	0.12	0.00
SaccadicSpeed	0.934	0.12	0.00
PredictiveSaccades	-0.063	0.21	0.00
BodyFixation	-0.007	0.21	0.00
ScanpathVariance	3.445	0.21	0.00
NonSpeakerSilhouetteCorrelation	-0.003	0.25	0.00
EyesFixation	0.006	0.52	0.00
NonSpeakerTimeCorrelation	-0.002	0.52	0.00
TimePerFixation	3698.224	0.73	0.00
FaceFixation	-0.003	0.88	0.00
ObjectFixation	0.000	0.96	0.00
ScreenTime	-0.001	0.96	0.00
SpeakerTimeCorrelation	-0.000	0.96	0.00

In summary, the gaze analysis features examined in this study were predominantly influenced by video

content and participant gender, with the effect of video content yielding substantially larger effect sizes. In contrast, participant age exhibited no impact on the extracted features. However, the limited age range within the dataset likely constrained the ability to detect age-related effects.

3.3. Statistical Analysis

Autism Diagnoses

All participants

The GEE-GLM analysis including all participants did not reveal any statistically significant associations between the eye-tracking features and autism diagnosis after correction for multiple comparisons, as detailed in Table 3.6. All adjusted p-values exceeded the 0.05 threshold, with effect sizes ranging from near-zero to 0.009, indicating low explanatory power for individual features in relation to diagnostic status. Corresponding plots are provided in Appendix F.

Table 3.6: GEE-GLM results for the relationship between each feature and autism diagnosis, ordered by increasing p-value. A positive coefficient suggests a higher feature value indicates a higher ASD probability.

Feature	Participants	Coefficient	p	η^2
EyesFixation	all	-4.60E-05	0.31	0.003
	male	-0.000324	0.56	0.003
SpeakerSilhouetteCorrelation	all	1.50E-05	0.35	0.002
	male	0.000176	0.56	0.005
PredictiveSaccades	all	7.00E-06	0.47	0.001
	male	-5.00E-05	0.56	0.002
ObjectFixation	all	1.70E-05	0.47	0.002
	male	0.000149	0.64	0.003
ScanpathVariance	all	1.00E-05	0.56	0.003
	male	0.000177	0.56	0.004
NonSpeakerTimeCorrelation	all	-5.00E-06	0.56	0.002
	male	-9.90E-05	0.56	0.004
FaceFixation	all	-2.10E-05	0.56	0.005
	male	-0.000263	0.56	0.004
ScreenTime	all	1.10E-05	0.74	0.000
	male	0.000143	0.85	0.000
BackgroundFixation	all	-6.00E-06	0.74	0.001
	male	8.90E-05	0.56	0.001
BodyFixation	all	1.20E-05	0.74	0.009
	male	0.000279	0.56	0.006
MouthFixation	all	8.00E-06	0.74	0.000
	male	6.50E-05	0.86	0.000
SaccadicSpeed	all	-2.00E-06	0.90	0.000
	male	0.000263	0.56	0.000
TimePerFixation	all	-3.00E-06	0.90	0.000
	male	4.50E-05	0.86	0.000
SpeakerTimeCorrelation	all	-3.00E-06	0.90	0.000
	male	2.60E-05	0.86	0.000
NonSpeakerSilhouetteCorrelation	all	1.00E-06	0.90	0.000
	male	-1.80E-05	0.86	0.000

Male-only model

Restricting GEE-GLM analyses for autism diagnosis to male participants, intended to address the known underdiagnosis of females, did not yield higher effect sizes or more statistically significant outcomes, as detailed in Table 3.6. This may be partly attributable to the associated reduction in sample size. Corresponding visualisations are presented in Appendix G.

SRS Scores

All participants

The GEE-GLM analysis including all participants identified five features that were statistically significantly associated with SRS scores after multiple comparison correction. These included Background Fixation, Speaker Time Correlation, Face Fixation, Predictive Saccades and Scanpath Variance, all showing adjusted p-values below 0.05 and effect sizes ranging from 0.002 to 0.004, as shown in Table 3.7. Effect sizes were small despite statistical significance, indicating subtle but reliable relationships. All variables were standardised using z-scores, so the reported coefficients represent the expected change in the outcome in standard deviation units for a one standard deviation change in the predictor. In contrast, ten features did not reach statistical significance. These features were Eye Fixation, Body Fixation, Screen Time, Mouth Fixation, Speaker Silhouette Correlation, Saccadic Speed, Non-Speaker Silhouette Correlation, Non-Speaker Time Correlation, Time Per Fixation and Object Fixation. Figures with SRS scores plotted against feature values are provided in Appendix H.

Table 3.7: GEE-GLM results for the relationship between each feature and SRS scores, ordered by increasing p-value. A positive coefficient suggests a higher feature value indicates a higher SRS score, a negative coefficient suggests a higher feature value indicates a lower SRS score.

Feature	**Coefficient**	p	η^2
BackgroundFixation	0.000113	0.00	0.004
SpeakerTimeCorrelation	-0.000121	0.01	0.003
FaceFixation	-0.000244	0.01	0.004
ScanpathVariance	0.000103	0.01	0.003
PredictiveSaccades	-5.95E-05	0.02	0.002
EyesFixation	-0.000194	0.05	0.002
BodyFixation	0.000132	0.06	0.002
ScreenTime	-0.000148	0.21	0.001
MouthFixation	-6.96E-05	0.38	0.000
SpeakerSilhouetteCorrelation	-2.94E-05	0.53	0.000
SaccadicSpeed	-7.69E-05	0.58	0.000
NonSpeakerSilhouetteCorrelation	-1.17E-05	0.56	0.000
NonSpeakerTimeCorrelation	1.20E-05	0.67	0.000
ObjectFixation	1.25E-05	0.83	0.000
TimePerFixation	-8.74E-06	0.90	0.000

Gender-Specific Patterns in Behaviours Indicative of Social Responsiveness

Analysis of the gender interaction terms revealed no features for which the relationship with SRS scores differed significantly by participant gender, as shown in Table 3.8. Therefore, no separate feature sets for male and female participants were created. Supplementary figures illustrating the relationships between each feature and SRS scores by gender are provided in Appendix I.

Table 3.8: P-values for differences in the relationship between feature and SRS score between genders, ordered by increasing p-value.

Predictor	p
ScanpathVariance	0.08
BackgroundFixation	0.08
PredictiveSaccades	0.10
SpeakerTimeCorrelation	0.10
FaceFixation	0.21
EyesFixation	0.45
ScreenTime	0.64
NonSpeakerTimeCorrelation	0.64
BodyFixation	0.64
MouthFixation	0.64
TimePerFixation	0.76
SaccadicSpeed	0.85
NonSpeakerSilhouetteCorrelation	0.85
ObjectFixation	0.85
SpeakerSilhouetteCorrelation	0.87

3.4. Predictive Modelling

Feature Sets

The feature sets were defined by including features that were either statistically significant ($p < 0.05$) or exhibited a partial effect size of $\eta^2 \geq 0.002$. Features that met either of these requirements in the male-only model for autism diagnosis but not in the all participants model were included, as females are known to be underdiagnosed, likely making the male-only model more accurate. The inclusion of non-significant features with effect sizes above this threshold was motivated by mathematical difficulty of achieving statistical significance in models with binary outcomes compared to those with continuous outcomes. [45] In the context of the continuous outcome of SRS scores, several features with $\eta^2 \geq 0.002$ were statistically significant, suggesting that this level of effect size may reflect meaningful associations. Therefore, excluding features on the basis of a p-value threshold might result in the omission of potentially informative predictors in the binary diagnostic analysis. The feature set for autism diagnosis consisted of following eight features: scanpath variance, speaker silhouette correlation, non-speaker time correlation, predictive saccades, body fixation, eye fixation, face fixation and object fixation. The feature set for SRS scores consisted of the following five features: speaker time correlation, predictive saccades, scanpath variance, background fixation and face fixation. Three features overlap between the feature sets predicting autism diagnosis and SRS scores.

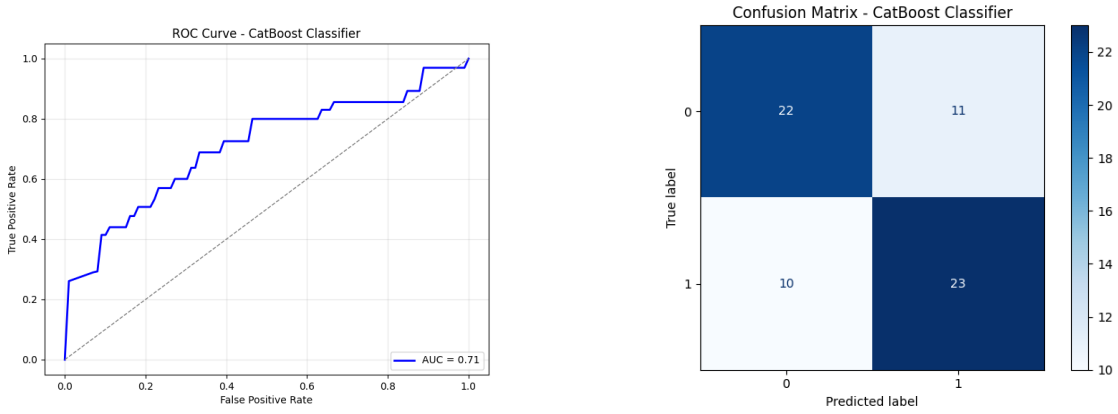
Feature Redundancy

Eye fixation was removed from the ASD feature set based on domain considerations specific to this study about feature redundancy: the eyes were part of the face in our data analysis. The decision to remove eyes fixation and keep face fixation was based on the higher effect size of face fixation. Correlation matrices and VIF values of all other features, presented in Appendix J, confirmed that multicollinearity was not an issue in either the ASD or SRS feature sets. The maximum correlation observed between any two features was 0.60 in the ASD set and 0.58 in the SRS set. Both values are well below the predefined threshold of 0.80, and indicate a moderate degree of linear relationship

between features. Furthermore, the maximum VIF was 2.87 for the ASD feature set and 1.93 for the SRS feature set. Since both values are considerably lower than the established threshold of 10, no additional features needed to be excluded due to redundancy. Therefore, the final feature set for ASD predictive modelling consisted of seven features, and the final feature set for SRS score predictive modelling consisted of five features.

ASD Model

Results of the hyperparameter tuning for all ASD models are provided in Appendix K. Among the evaluated models, the CatBoost Gradient Boosting Decision Tree emerged as the most effective approach for ASD modelling, with an AUC of 0.71. Its corresponding ROC curve and confusion matrix are shown in Figure 3.1. The F1 score this model achieved was 0.69, with a precision of 0.69 and a recall of 0.72.



(a) ROC curve

(b) Confusion matrix

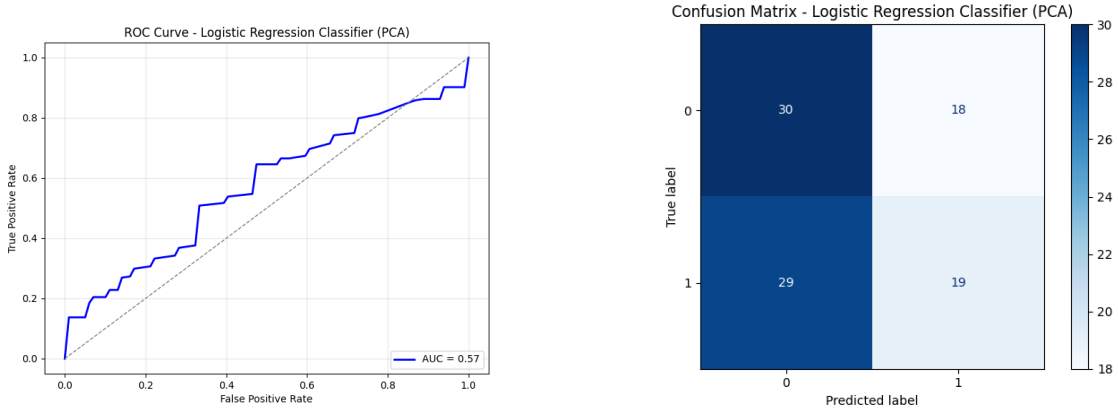
Figure 3.1: Evaluation metrics for the CatBoost Gradient Boosting Decision Tree classifier: AUC = 0.71; F1 = 0.69; precision = 0.69; recall = 0.72.

The AUC of 0.71, considering the feature distributions and small effect sizes, demonstrates how combining multiple subtle signals in a powerful model can create a reasonably discriminative classifier. Performance metrics for all other ASD classifiers, including ROC curves, confusion matrices, F1 scores, precision, and recall, are presented in Appendix L. Higher AUC scores were achieved with increased model complexity. This suggests that the data contains non-linear and high-dimensional structure that cannot be adequately captured by linear models. Neither PCA nor LDA significantly improved the performance of the Logistic Regression model. Similarly, ensemble methods did not outperform the best-performing standalone model.

SRS Model

Results of the hyperparameter tuning for all SRS models are provided in Appendix K. Although the more complex individual models outperformed Logistic Regression, the highest performance was achieved by the Logistic Regression Classifier combined with PCA, leading to an AUC of 0.57. Its ROC curve and confusion matrix are provided in Figure M.4. This classifier attained an F1 score of 0.42, a precision of 0.65 and a recall of 0.41.

Both dimensionality reduction techniques increased performance, which indicates the raw features for SRS are noisy. Additionally, PCA outperformed LDA, which may indicate that class-discriminative information in the data is not well captured by linear class separation, possibly reflecting underlying nonlinear structure. Performance metrics for all other SRS classifiers, including ROC curves, confusion matrices, F1 scores, precision, and recall, are presented in Appendix M.



(a) ROC curve

(b) Confusion matrix

Figure 3.2: Evaluation metrics for the Logistic Regression classifier with PCA: AUC = 0.57; F1 = 0.42; precision = 0.65; recall = 0.41.

4

Discussion

4.1. Conclusions

This study examined gaze patterns related to ASD and social responsiveness, revealing interactions that underscore the potential of gaze analysis for both understanding and identifying ASD. A predictive model for ASD achieved an AUC of 0.71, indicating that gaze-based features capture meaningful differences in social information processing between autistic and neurotypical individuals. Unexpectedly, the predictive model for SRS scores achieved a lower AUC of 0.57, despite consistent SRS evaluation across all participants, whereas ASD diagnosis relied on self-initiated clinical referral. A possible explanation for the difference in performance is that social responsiveness, as assessed by the SRS, represents a component of ASD that is more amenable to social learning and compensation, particularly during adolescence. Another possibility is that certain non-social characteristics of autism, such as motor control differences, which affect up to 87% of autistic individuals, may be reflected in features like predictive saccades and scanpath variance. [46] These aspects may thus offer a more stable and less consciously masked signal for identifying ASD. Collectively, these findings underscore both the promise and challenges of applying gaze-based metrics in autism research. While eye tracking shows potential for ASD detection, substantial methodological and conceptual work remains before gaze analysis can be applied in clinical practice. To situate these findings in context, the following section reviews how they compare with prior work on ASD detection using eye tracking.

4.2. Comparison with Existing Literature

The majority of studies included in the literature review reported only the presence or absence of a correlation with the outcome measure, without providing predictive performance metrics. However, some studies did report predictive performance metrics. Notably, Constantino et al. achieved an AUC of 0.86 in infants using fixation duration features focused on the eyes and mouth. [22] This relatively high AUC is likely attributable to the more appropriate age range of the participants, as infants have not yet developed socially desirable behaviors that mask their natural gaze behaviour. Müller et al. reported an AUC of 0.75 by combining visual attention to the eyes, mouth, body, and objects with pupil dilation, a feature found to be moderately reduced in individuals with ASD. [47] Similarly, a large-scale study in toddlers by Chang et al., which utilised gaze-speech synchrony as the primary feature, yielded an AUC of 0.76. [48] Cilia et al. achieved an AUC of 0.71 using scanpath images, while Liaqat et al. who also

focused on scanpaths reported a lower AUC of 0.64. [49, 50] In summary, the findings of this study align with existing literature. Higher predictive performance is generally observed in younger populations, such as infants and toddlers, underscoring the likelihood of masking effects in older children and adolescents. However, an important factor to consider before reaching this conclusion is that participants included at a young age based on an ASD diagnosis are likely more severe cases. The ideal approach would involve conducting an eye-tracking task at a young age and longitudinally tracking the cohort over several years to identify who eventually receives an ASD diagnosis. However, this method requires significantly more time and financial resources.

4.3. Applicability and Generalisability

These patterns of age-related variation in performance raise important questions about how well eye-tracking tools generalise across different populations and settings. Applicability and generalisability are important for scalability, and therefore for the clinical utility of diagnostic tools. Applicability and generalisability can be studies from two key perspectives: first, whether the tool performs consistently across diverse population groups, and second, whether it remains effective when applied with different video stimuli.

First of all, the consistency of gaze-based features across demographic variables such as culture, age, and gender must be assessed. Previous research has demonstrated the cross-cultural validity of eye-tracking parameters in the detection of ASD. [23, 24] In contrast, gender does require correction of gaze-derived features, as confirmed by our data analysis. [24] While limited information is available about the performance of similar eye-tracking tools, existing studies suggest better performance in toddlers and infants than in adolescents or adults. [22, 47, 48] Taken together, these findings suggest that gaze-based features are broadly applicable across cultures, especially in younger populations, provided that gender-related differences are accounted for.

Regarding the applicability of different video stimuli, important considerations must be taken into account. Gepner et al. demonstrated that a reduced pace of speech increases visual attention to the speaker's mouth in individuals with ASD. They propose that slower speech allows more time for processing facial dynamics, which may be cognitively overwhelming at typical conversational speeds for individuals with ASD. [51] As a consequence, cut-off values for certain features, such as face fixation duration, may vary depending on the speech tempo. This suggests that calibration on groups with and without ASD remained essential prior to clinical application of new video stimuli. In addition to speech pace, the emotional content of a video can also modulate gaze behaviour. Research by Sahuquillo-Leal et al. has shown that children with ASD exhibit more avoidant gaze behaviour in response to threatening scenes compared to TD peers. [52] These findings underscore that both linguistic and emotional characteristics of stimuli can substantially influence visual attention patterns.

Taken together, these results indicate that both participant characteristics and stimulus characteristics play critical roles in shaping gaze behaviour. Consequently, deliberate stimulus design choices are essential to maximising the diagnostic utility of gaze-based tools.

4.4. Design for Detection: Stimulus Design

Building on this need for deliberate design, the following section explores specific stimulus characteristics that can enhance sensitivity to ASD-related gaze behaviour. As explained, video recordings of naturalistic social interactions involving multiple individuals appear to be the most effective stimulus type for identifying gaze patterns associated with ASD. Building on this foundation, several additional features may further enhance the discriminative power of gaze-based assessments and should be con-

sidered in future stimulus design. Future stimuli could be refined to systematically elicit these features.

Scene inversion, achieved by mirroring the scene, represents one such feature that has not been utilised in the current set of stimuli. Shic et al. demonstrated that children with ASD are less affected by scene inversion compared to TD children, suggesting reduced sensitivity to configural social information. [53] Another promising feature is gaze idiosyncrasy, which can be captured by presenting the same video stimulus multiple times. It has been shown that children with ASD observe social interactions in a highly idiosyncratic manner. This manifests not only between individuals, as computed with the scanpath variance feature, but also within individuals across repeated viewings. While TD children have highly correlated scanpaths across multiple viewings, children with ASD show significantly lower within-subject consistency. [54] Another candidate feature is eye blink synchronisation. Nakano et al. reported that TD adults tend to synchronise their eye blinks with a speaker, reflecting attentional engagement and social attunement. In contrast, adults with ASD do not show this synchronisation. [55] Moreover, frame-by-frame analysis has yet to be utilised in this context, but holds potential to enhance classification performance by capturing moment-to-moment variations in gaze behaviour. Incorporating a non-verbal request in the stimulus results in reduced fixation on the person receiving that request in individuals with ASD and could therefore enhance classification performance. [56, 57] Finally, the video stimuli used in this study could be improved by using a geometric image as the distracting object alongside the video, rather than a poster in the background, as children with ASD have a strong preference for geometric images. [58–61] De Belen et al. achieved an AUC of 0.96 using this method with images.

In summary, multiple gaze-derived features hold meaningful potential to improve design of future stimuli for the detection of ASD-related visual behaviour. Incorporating both regular and inverted versions of the same scenes, presenting scenes multiple times, assessing eye blink synchronisation with the character onscreen and including a geometric image could significantly enhance the discriminative power of the stimuli.

4.5. Multimodal Extension Opportunities

While optimising visual stimuli is one strategy for improving detection, integrating additional physiological modalities may further enhance classification performance. In their 2019 study, Dijkhuis et al. demonstrated that adults with ASD exhibit a lower skin conductance level (SCL) when viewing socio-emotional video clips, suggesting that SCL may serve as a valuable complementary feature in ASD detection within the same experimental design. [62] In addition, electroencephalography (EEG) offers potential to enhance diagnostic accuracy. Specifically, differences in visual interest in biological motion between children with ASD and TD peers can be observed through reduced suppression of the alpha and high beta frequency bands in the right frontal and parietal regions. [63] When EEG is employed, its utility can be extended by incorporating motor-related stimuli. TD individuals typically show mu-band suppression both when performing and observing movement, whereas individuals with ASD display mu-suppression only during execution of movement, not during observation. [64] A potential drawback of incorporating physiological measures such as SCL and EEG is the associated increase in cost and technical complexity. Nevertheless, recent developments offer promising solutions in terms of cost-effectiveness. For instance, Vargas-Cuentas et al. developed a portable tablet-based setup that eliminates the need for eye-tracker calibration or head stabilisation, thereby significantly reducing hardware costs while maintaining adequate accuracy for gaze-based assessment. [65] In summary, while the integration of multimodal measures such as SCL and EEG can substantially enhance predictive performance, recent technological advances also offer pathways to achieve this in a cost-effective manner.

In additional to technical modalitites, an extension of the algorithm to include patient characteristics like family history of autism is also a possible direction of improvements.

4.6. General Strengths and Limitations

With these technical and conceptual opportunities in mind, it is important to consider the strengths and limitations of the current study in order to contextualise its contributions. First of all, this study possesses several notable strengths. The use of naturalistic, dynamic video stimuli ensures high ecological validity, providing a more realistic context for assessing visual attention compared to static or artificial stimuli. Additionally, the multifaceted set of gaze features enables comprehensive modelling of different aspects of visual behaviour, from saccadic dynamics to fixation patterns. The inclusion of both categorical ASD diagnosis and dimensional SRS scores as outcome variables permits an examination of autistic traits across different measurement frameworks. Furthermore, the comparative evaluation of multiple predictive modelling approaches offers insight into the data's underlying structure.

However, several limitations must also be acknowledged. First and foremost, the study is limited by the small sample size of participants with ASD and elevated SRS scores. Although the Generation R cohort represents a large-scale population database, it was not specifically designed to recruit substantial numbers of individuals with autism or pronounced autistic traits, constraining statistical power. Moreover, the study relies on ASD diagnoses obtained by individuals who independently sought help, which introduces bias and reduces the generalisability of the findings. It is possible that certain subgroups, particularly females and individuals who are high-functioning or engage in masking behaviours, are therefore not adequately labelled in the data. Additionally, the SRS, especially in its shortened form, does not fully capture the complexity of the autism phenotype. This lack of a ground truth for ASD diagnosis within the dataset complicates the validity of the results. Furthermore, the age of the participants proposes a methodological limitation. At approximately thirteen years old, they may have developed compensatory strategies or masking behaviours, potentially obscuring the natural behaviour. Ideally, eye-tracking studies should be conducted with children under the age of four, a period during which current diagnostic tools face challenges due to developmental variability, and where eye-tracking may be advantageous, as younger children are less likely to exhibit socially conditioned masking behaviours. Another limitation relates to the lack of control for comorbid factors such as ADHD or cognitive ability. Both of these factors are known to influence SRS scores and social attention but were not accounted for in the current analyses. [66, 67] These strengths and limitations highlight critical areas for future research, including improving diagnostic ground truths and accounting for co-occurring factors.

4.7. Data Analysis Limitations

In addition to general methodological constraints, there are limitations related specifically to the data analysis procedures employed in this study. Firstly, although the preprocessing steps were essential to ensure data quality, they also reduced the volume of usable data and may have introduced bias. For example, screen time was used both as a feature and as a quality control criterion, with participants excluded if their screen time fell below a 0.25 threshold. This may have biased the sample by excluding individuals with low screen time, who may represent individuals with limited social engagement. Additionally, the feature engineering process leads to loss of temporal granularity. By aggregating behavior across time windows, the analysis may overlook short-term fluctuations in gaze behavior that could be indicative of social processing. Together, these constraints point to a number of concrete avenues for improvement.

4.8. Recommendations

The following recommendations aim to guide future research design. To begin with, several strategies could be employed to explore alternative machine learning models. Using alternative forms of input data may unlock latent value within the dataset. One such approach is to represent the scanpath on a two-dimensional image and use this image as the input for a machine learning model. A convolutional neural network would be a suitable model architecture for this task, as these are well-established for processing visual input. This approach was carried out by Cilia et al., leading to an AUC of 0.71 after correcting the data leakage mistake in their machine learning approach. [49] Applying this technique to our dataset was not feasible as their were eight different videos, and unlike numeric feature values, scanpath images cannot be corrected for varying stimuli. Consequently, only measurements using the same stimulus can be compared, which results in a sample size too small to support meaningful machine learning analysis. Nonetheless, the MATLAB code used to generate the scanpath images is available for use.

Alternatively, it may be beneficial to use the raw gaze data, including fixation, saccade and blink registrations, directly as input. This strategy could uncover deeper structures within the data that may have been obscured or lost through extraction of predefined features. Colonnese et al. achieved an F1-score of 93% using raw gaze data and a variety of deep neural networks. [68] Important to note is that using the raw gaze data also requires splitting the dataset by video stimuli, which reduces the sample size to the same extent as the scanpath images method. Another major drawback of both these methods is their increasing loss of explainability, which is highly valuable in clinical research.

To enhance the predictive power of the machine learning models in this study, extracting additional features may prove beneficial. Head movements have been shown to be atypical in response to a social stimulus in individuals with ASD, making it a potentially valuable feature. [69] Another promising feature to incorporate in future studies is pupil dilation, which can be detected during video viewing and yields predictive value for ASD. [47] Furthermore, various other features could be derived when stimuli are optimised as described in Section 4.4.

In terms of study population of future studies, a much younger population under the age of four would be ideal. During this stage, current diagnostic tools face major challenges due to developmental variability, whilst eye-tracking may be advantageous as younger children are less likely to exhibit masking behaviours.

Regarding outcome measures, future analyses that have access to datasets that include the full version of the SRS, could benefit from examining SRS subscale scores. Unlike the short form, the full version of the Social Responsiveness Scale includes five distinct subdomains: social awareness, social cognition, social communication, social motivation, and restricted and repetitive behaviours. Investigating associations between these subscales and gaze-derived features may yield more nuanced insights into the specific aspects of social functioning reflected in gaze patterns.

These recommendations cover analytical techniques, stimulus refinement, population targeting, and outcome measure selection, together forming a roadmap for optimising future applications of gaze-based ASD detection. In conclusion, gaze behaviour holds considerable promise for generating deeper insights into ASD. These insights ultimately contribute to the development of low-cost, inclusive, and scalable ASD assessment tools, thereby enhancing both individual quality of life and broader socio-economic outcomes across diverse populations.

Bibliography

1. Zeidan, J. *et al.* Global prevalence of autism: A systematic review update. en. *Autism Res.* **15,** 778–790 (May 2022).
2. Nadeem, M. S. *et al.* Autism - A comprehensive array of prominent signs and symptoms. en. *Curr. Pharm. Des.* **27,** 1418–1433 (2021).
3. Park, S. H. *et al.* Disability, functioning, and quality of life among treatment-seeking young autistic adults and its relation to depression, anxiety, and stress. en. *Autism* **23,** 1675–1686 (Oct. 2019).
4. Movsessian, T. & Osoba, T. A. Association between therapeutic interventions and quality of life in people with autism. *J. Soc. Behav. Health Sci.* **16** (Nov. 2022).
5. Tedla, J. S. *et al.* Assessing the quality of life in children with autism spectrum disorder: a cross-sectional study of contributing factors. en. *Front. Psychiatry* **15,** 1507856 (Dec. 2024).
6. Bishop, S. L. & Lord, C. Commentary: Best practices and processes for assessment of autism spectrum disorder - the intended role of standardized diagnostic instruments. en. *J. Child Psychol. Psychiatry* **64,** 834–838 (May 2023).
7. Huda, E. *et al.* Screening tools for autism in culturally and linguistically diverse paediatric populations: a systematic review. en. *BMC Pediatr.* **24,** 610 (Sept. 2024).
8. Lefort-Besnard, J. *et al.* Patterns of autism symptoms: hidden structure in the ADOS and ADI-R instruments. en. *Transl. Psychiatry* **10,** 257 (July 2020).
9. Hus, Y. & Segal, O. Challenges surrounding the diagnosis of autism in children. en. *Neuropsychiatr. Dis. Treat.* **17,** 3509–3529 (Dec. 2021).
10. Okoye, C. *et al.* Early diagnosis of autism spectrum disorder: A review and analysis of the risks and benefits. en. *Cureus* **15,** e43226 (Aug. 2023).
11. Estes, A. *et al.* Long-term outcomes of early intervention in 6-year-old children with autism spectrum disorder. en. *J. Am. Acad. Child Adolesc. Psychiatry* **54,** 580–587 (July 2015).
12. Vivanti, G., Prior, M., Williams, K. & Dissanayake, C. Predictors of outcomes in autism early intervention: Why don't we know more? en. *Front. Pediatr.* **2,** 58 (June 2014).
13. Andrade, C. Autism spectrum disorder, 1: Genetic and environmental risk factors. en. *J. Clin. Psychiatry* **86** (Apr. 2025).
14. Strathearn, L. The elusive etiology of autism: nature and nurture? en. *Front. Behav. Neurosci.* **3,** 11 (July 2009).
15. Matuskey, D. *et al.* 11C-UCB-J PET imaging is consistent with lower synaptic density in autistic adults. en. *Mol. Psychiatry* (Oct. 2024).
16. Gandal, M. J. *et al.* Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. en. *Nature* **611,** 532–539 (Nov. 2022).
17. Papagiannopoulou, E. A., Chitty, K. M., Hermens, D. F., Hickie, I. B. & Lagopoulos, J. A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. en. *Soc. Neurosci.* **9,** 610–632 (July 2014).

18. Pelphrey, K. A., Morris, J. P. & McCarthy, G. Neural basis of eye gaze processing deficits in autism. en. *Brain* **128,** 1038–1048 (May 2005).

19. Agam, Y., Joseph, R. M., Barton, J. J. S. & Manoach, D. S. Reduced cognitive control of response inhibition by the anterior cingulate cortex in autism spectrum disorders. en. *Neuroimage* **52,** 336–347 (Aug. 2010).

20. Bölte, S. *et al.* How can clinicians detect and treat autism early? Methodological trends of technology use in research. en. *Acta Paediatr.* **105,** 137–144 (Feb. 2016).

21. Klin, A. Biomarkers in autism spectrum disorder: Challenges, advances, and the need for biomarkers of relevance to public health. en. *Focus (Am. Psychiatr. Publ.)* **16,** 135–142 (Apr. 2018).

22. Constantino, J. N. *et al.* Infant viewing of social scenes is under genetic control and is atypical in autism. en. *Nature* **547,** 340–344 (July 2017).

23. Al-Shaban, F. A. *et al.* Development and validation of an Arabic language eye-tracking paradigm for the early screening and diagnosis of autism spectrum disorders in Qatar. en. *Autism Res.* **16,** 2291–2301 (Dec. 2023).

24. Frazier, T. W. *et al.* Social attention as a cross-cultural transdiagnostic neurodevelopmental risk marker. en. *Autism Res.* **14,** 1873–1885 (Sept. 2021).

25. Speer, L. L., Cook, A. E., McMahon, W. M. & Clark, E. Face processing in children with autism: effects of stimulus contents and type. en. *Autism* **11,** 265–277 (May 2007).

26. Chita-Tegmark, M. Social attention in ASD: A review and meta-analysis of eye-tracking studies. en. *Res. Dev. Disabil.* **48,** 79–93 (Jan. 2016).

27. Grossman, R. B., Zane, E., Mertens, J. & Mitchell, T. Facetime vs. Screentime: Gaze patterns to live and video social stimuli in adolescents with ASD. en. *Sci. Rep.* **9,** 12643 (Sept. 2019).

28. Robain, F. *et al.* Measuring social orienting in preschoolers with autism spectrum disorder using cartoons stimuli. en. *J. Psychiatr. Res.* **156,** 398–405 (Dec. 2022).

29. Parker, T. C. *et al.* Neural and visual processing of social gaze cueing in typical and ASD adults. en. *medRxiv* (Feb. 2023).

30. Bezemer, M. L., Blijd-Hoogewys, E. M. A. & Meek-Heekelaar, M. The predictive value of the AQ and the SRS-A in the diagnosis of ASD in adults in clinical practice. en. *J. Autism Dev. Disord.* **51,** 2402–2415 (July 2021).

31. Mooldijk, S. S., Labrecque, J. A., Ikram, M. A. & Ikram, M. K. Ratios in regression analyses with causal questions. en. *Am. J. Epidemiol.* **194,** 311–313 (Jan. 2025).

32. Engel, K. C., Anderson, J. H. & Soechting, J. F. Oculomotor tracking in two dimensions. *J. Neurophysiol.* **81,** 1597–1602 (1999).

33. Bahill, A. T., Clark, M. R. & Stark, L. The main sequence, a tool for studying human eye movements. en. *Math. Biosci.* **24,** 191–204 (Jan. 1975).

34. Hessels, R. S., Kemner, C., van den Boomen, C. & Hooge, I. T. C. The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli. en. *Behav. Res. Methods* **48,** 1694–1712 (Dec. 2016).

35. Frank, M. C., Vul, E. & Saxe, R. Measuring the development of social attention using free-viewing. en. *Infancy* **17,** 355–375 (July 2012).

36. Orquin, J. L., Ashby, N. J. S. & Clarke, A. D. F. Areas of interest as a signal detection problem in behavioral eye tracking research. en. *J. Behav. Decis. Mak.* **29,** 103–115 (Apr. 2016).

37. Hus, V., Bishop, S., Gotham, K., Huerta, M. & Lord, C. Factors influencing scores on the social responsiveness scale. en. *J. Child Psychol. Psychiatry* **54,** 216–224 (Feb. 2013).

38. Lockwood Estrin, G., Milner, V., Spain, D., Happé, F. & Colvert, E. Barriers to autism spectrum disorder diagnosis for young women and girls: A systematic review. en. *Rev. J. Autism Dev. Disord.* **8,** 454–470 (2021).

39. Halladay, A. K. *et al.* Sex and gender differences in autism spectrum disorder: summarizing evidence gaps and identifying emerging areas of priority. en. *Mol. Autism* **6,** 36 (June 2015).

40. Wood-Downie, H., Wong, B., Kovshoff, H., Cortese, S. & Hadwin, J. A. Research Review: A systematic review and meta-analysis of sex/gender differences in social interaction and communication in autistic and nonautistic children and adolescents. en. *J. Child Psychol. Psychiatry* **62,** 922–936 (Aug. 2021).

41. Lundin, K., Mahdi, S., Isaksson, J. & Bölte, S. Functional gender differences in autism: An international, multidisciplinary expert survey using the International Classification of Functioning, Disability, and Health model. en. *Autism* **25,** 1020–1035 (May 2021).

42. Pennsylvania State University. *10.7 - Detecting Multicollinearity Using Variance Inflation Factors* Accessed: April 11, 2025. 2025. https://online.stat.psu.edu/stat462/node/180/.

43. Nguyen, P. H. *et al.* The reliability and validity of the social responsiveness scale to measure autism symptomology in Vietnamese children. en. *Autism Res.* **12,** 1706–1718 (Nov. 2019).

44. Lyall, K. *et al.* Examining shortened versions of the Social Responsiveness Scale for use in autism spectrum disorder prediction and as a quantitative trait measure: Results from a validation study of 3-5 year old children. en. *JCPP Adv.* **2,** e12106 (Dec. 2022).

45. Suissa, S. & Blais, L. Binary regression with continuous outcomes. en. *Stat. Med.* **14,** 247–255 (Feb. 1995).

46. Bhat, A. N. Motor impairment increases in children with autism spectrum disorder as a function of social communication, cognitive and functional impairment, repetitive behavior severity, and comorbid diagnoses: A SPARK study report. en. *Autism Res.* **14,** 202–219 (Jan. 2021).

47. Müller, N., Baumeister, S., Dziobek, I., Banaschewski, T. & Poustka, L. Validation of the Movie for the Assessment of Social Cognition in Adolescents with ASD: Fixation Duration and Pupil Dilation as Predictors of Performance. *Journal of Autism and Developmental Disorders* **46,** 2831–2844. ISSN: 1573-3432. http://dx.doi.org/10.1007/s10803-016-2828-z (June 2016).

48. Chang, Z. *et al.* Computational Methods to Measure Patterns of Gaze in Toddlers With Autism Spectrum Disorder. *JAMA Pediatrics.* ISSN: 2168-6203. http://dx.doi.org/10.1001/jamapediatrics.2021.0530 (Apr. 2021).

49. Cilia, F. *et al.* Computer-aided screening of autism spectrum disorder: Eye-tracking study using data visualization and deep learning. en. *JMIR Hum. Factors* **8,** e27706 (Oct. 2021).

50. Liaqat, S. *et al.* Predicting ASD diagnosis in children with synthetic and image-based eye gaze data. en. *Signal Process. Image Commun.* **94,** 116198 (May 2021).

51. Gepner, B., Godde, A., Charrier, A., Carvalho, N. & Tardif, C. Reducing facial dynamics' speed during speech enhances attention to mouth in children with autism spectrum disorder: An eye-tracking study. en. *Dev. Psychopathol.* **33,** 1006–1015 (Aug. 2021).

52. Sahuquillo-Leal, R. *et al.* Attentional biases towards emotional scenes in autism spectrum condition: An eye-tracking study. en. *Res. Dev. Disabil.* **120,** 104124 (Jan. 2022).

53. Shic, F., Scassellati, B., Lin, D. & Chawarska, K. *Measuring context: The gaze patterns of children with autism evaluated from the bottom-up* in *2007 IEEE 6th International Conference on Development and Learning* (IEEE, London, UK, July 2007).

54. Avni, I. *et al.* Children with autism observe social interactions in an idiosyncratic manner. en. *Autism Res.* **13,** 935–946 (June 2020).

55. Nakano, T., Kato, N. & Kitazawa, S. Lack of eyeblink entrainments in autism spectrum disorders. en. *Neuropsychologia* **49,** 2784–2790 (July 2011).

56. Falck-Ytter, T., von Hofsten, C., Gillberg, C. & Fernell, E. Visualization and analysis of eye movement data from children with typical and atypical development. en. *J. Autism Dev. Disord.* **43,** 2249–2258 (Oct. 2013).

57. Viktorsson, C., Bölte, S. & Falck-Ytter, T. How 18-month-olds with later autism look at other children interacting: The timing of gaze allocation. en. *J. Autism Dev. Disord.* **54,** 4091–4101 (Nov. 2024).

58. Pierce, K., Conant, D., Hazin, R., Stoner, R. & Desmond, J. Preference for geometric patterns early in life as a risk factor for autism. en. *Arch. Gen. Psychiatry* **68,** 101–109 (Jan. 2011).

59. Pierce, K. *et al.* Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder subtype associated with increased symptom severity. en. *Biol. Psychiatry* **79,** 657–666 (Apr. 2016).

60. Moore, A. *et al.* The geometric preference subtype in ASD: identifying a consistent, early-emerging phenomenon through eye tracking. en. *Mol. Autism* **9** (Dec. 2018).

61. De Belen, R. A. J., Eapen, V., Bednarz, T. & Sowmya, A. Using visual attention estimation on videos for automated prediction of autism spectrum disorder and symptom severity in preschool children. en. *PLoS One* **19,** e0282818 (Feb. 2024).

62. Dijkhuis, R., Gurbuz, E., Ziermans, T., Staal, W. & Swaab, H. Social attention and emotional responsiveness in young adults with autism. en. *Front. Psychiatry* **10,** 426 (June 2019).

63. Cantonas, L.-M. *et al. Impaired alpha and beta modulation in response to social stimuli in children with autism spectrum disorder* Jan. 2022.

64. Oberman, L. M. *et al.* EEG evidence for mirror neuron dysfunction in autism spectrum disorders. en. *Brain Res. Cogn. Brain Res.* **24,** 190–198 (July 2005).

65. Vargas-Cuentas, N. I. *et al.* Developing an eye-tracking algorithm as a potential tool for early diagnosis of autism spectrum disorder in children. en. *PLoS One* **12,** e0188826 (Nov. 2017).

66. Tian, J. *et al. Atypical biological motion perception in children with attention deficit hyperactivity disorder: Local motion and global configuration processing* May 2024.

67. Ayyildiz, D., Bikmazer, A., Cahid Örengül, A. & Perdahlı Fiş, N. Executive Functions and social responsiveness in children and adolescents with autism spectrum disorder and attention deficit hyperactivity disorder. en. *Psyc.. Clin. Psychopharmacol.* **31,** 165–172 (June 2021).

68. Colonnese, F., Di Luzio, F., Rosato, A. & Panella, M. Enhancing autism detection through gaze analysis using eye tracking sensors and data attribution with distillation in deep neural networks. en. *Sensors (Basel)* **24** (Dec. 2024).

69. Martin, K. B. *et al.* Objective measurement of head movement differences in children with and without autism spectrum disorder. en. *Mol. Autism* **9** (Dec. 2018).

A

Literature Review

Eye Tracking as a Biomarker for ASD: A Review of Eye Movement Outcome Parameters for Dynamic Social Stimuli

I.M. Brugman[1], R. van der Vliet[2], M. Frens[2]

February 26, 2025

[1]*Delft University of Technology*
[2]*Erasmus Medical Center*

Introduction: Autism Spectrum Disorder (ASD) diminishes quality of life, especially without early intervention. While current diagnostic tools facing limitations, eye tracking has emerged as an objective biomarker for ASD assessment. This review examines existing research on eye tracking outcome parameters when using naturalistic social interaction videos as a social stimulus.
Methods: A systematic search of MEDLINE, Embase, and Web of Science was conducted to identify empirical studies comparing eye movements in ASD and typically developing (TD) individuals while watching naturalistic social scene videos.
Results: Out of 1520 articles, 19 met the inclusion criteria, in which 13 outcome parameters were identified including areas of interest, gaze-speech correlation, scanpatch variance saccadic speed, total looking time, time per fixation and frame-by-frame analysis.
Conclusion: Eye-tracking is a promising biomarker for ASD, as individuals with ASD showed reduced fixation on social cues, increased attention to non-social elements, and altered gaze patterns.

1 Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that affects approximately 1 in 100 children worldwide. [1] The brains of individuals with ASD show significant differences compared to those of their typically developing (TD) peers. [2, 3] Core characteristics of ASD include deficits in social interaction, restrictive interests, anxiety, and repetitive behaviors. [4] Research indicates that young autistic adults experience anxiety, depression, and stress levels comparable to individuals diagnosed with primary anxiety and depressive disorders. [5] Moreover, the quality of life for individuals with ASD is consistently lower than that of the general population, across all age groups. [6, 7]

Early identification allows for timely intervention, which has been shown to significantly enhance developmental outcomes and quality of life for individuals with ASD. In addition to individual benefits, early intervention also yields significant economic benefits over time. [8, 9] Currently, gold-standard diagnostic instruments such as the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview (ADI) require extensive time for administration and analysis. Additionally, these tools are susceptible to subjectivity, thereby reducing diagnostic accuracy, particularly among certain subgroups. [10] For example, culturally and linguistically diverse populations often experience diminished diagnostic precision with these assessments. [11] Additionally, these tools frequently fail to adequately detect high-functioning autism, as they rely heavily on overt behavioural markers that may not capture the subtle social deficits of this subgroup. [12] Another limitation is their reduced applicability in very young children due to reliance on caregiver recall and the substantial developmental variability observed in children under the age of three. [13] Consequently, there is an urgent need for objective biomarkers in ASD assessment.

Eye tracking has emerged as a promising biomarker capable of differentiating ASD from TD individuals. Atypical gaze patterns in individuals with ASD are linked to dysfunctions across multiple brain regions, including the amygdala, frontal eye fields, temporal parietal junction, insula, and dorsal lateral prefrontal cortex. [14] Integrating eye tracking into the diagnostic process has been shown to enhance diagnostic efficiency while reducing the lifetime cost of autism. [15] Additionally, eye tracking mitigates the limitations associated with traditional diagnostic tools. It is particularly advantageous for studying infants and young children's cognitive processes in a non-intrusive manner. [16, 17] Furthermore, eye tracking holds promise as a solution for assessing high-functioning autism and culturally diverse populations due to the increasing availability of varied social stimuli

Previous studies suggest that dynamic social stimuli involving more than one person are optimal for distinguishing between ASD and TD individuals. [18, 19] The differences between these groups become more pronounced when video-based stimuli are used instead of live social interactions. [20] Moreover, naturalistic scenes yield greater ecological validity than animated characters. [21, 22] Taken together, it becomes evident that the most effective method for detecting ASD-related gaze patterns is through video recordings of naturalistic social interactions involving multiple individuals. However, the optimal outcome parameter for such dynamic stimuli remains undetermined, as previous studies predominantly employed static stimuli. Therefore, the objective of this review is to synthesize existing research on eye tracking outcome parameters during naturalistic dynamic social stimuli involving multiple individuals, with the goal of improving the diagnostic accuracy of ASD.

2 Methods

2.1 Search Strategy

The databases MEDLINE, Embase and Web of Science were searched up to February 3, 2025 to identify relevant articles. The search strategy was developed in collaboration with a medical librarian from the Erasmus Medical Center, using keywords such as "autism", "social stimulus", and "eye tracking". The complete search strategy is provided in Appendix A.

2.2 Inclusion Criteria

Inclusion criteria were (a) Published in English; (b) Empiral study; (c) Human participants; (d) Patient group of participants with autism-related traits (either defined as a diagnosis on the autism spectrum, or evaluated with social cognition scales, e.g. Social Responsiveness Scale (SRS), Autism Diagnostic Interview (ADI) or Autism Diagnostic Observation Schedule (ADOS)); (e) Control group of neurotypically developed peers; (f) The main goal of the study was to compare the eye movements of both groups during a social stimulus; (g) The social stimulus was a video of a naturalistic social scene involving multiple characters. When a study employed this type of stimulus along with others, the article was only included if a distinction in outcome parameters was made.

3 Results

3.1 Search Results

The PRISMA flow diagram of the search strategy is presented in Figure 1. The initial search yielded 1520 articles, with 715 remaining after duplicate removal. Title and abstract screening led to the exclusion of 668 studies. A full text review of the remaining 47 articles resulted in the inclusion of 19 studies in the final review..

3.2 Study Characteristics

The median sample size of the 19 included studies was 63 participants. The age range varied across studies, spanning from infants to adults. Some studies performed a between-group comparison of eye movements including both an ASD and TD group, while others examined the correlation between gaze behaviour and autism-related scales. The autism-related scales utilised in the included articles were the Autism Diagnostic Observation Schedule (ADOS), Social Responsiveness Scale (SRS), Multidimensional Anxiety Scale for Children (MASC), and the Vineland Adaptive Behaviour Scales Expanded edition (VABS-E).

3.3 Outcome parameters

The 13 identified outcome parameters consisted of seven Areas of Interest (AOI) and six other outcome parameters. An overview of the outcome parameters is provided in 1. Each of them will be discussed below. When a study combined eyes and mouth as a single outcome parameter, it was considered the face AOI

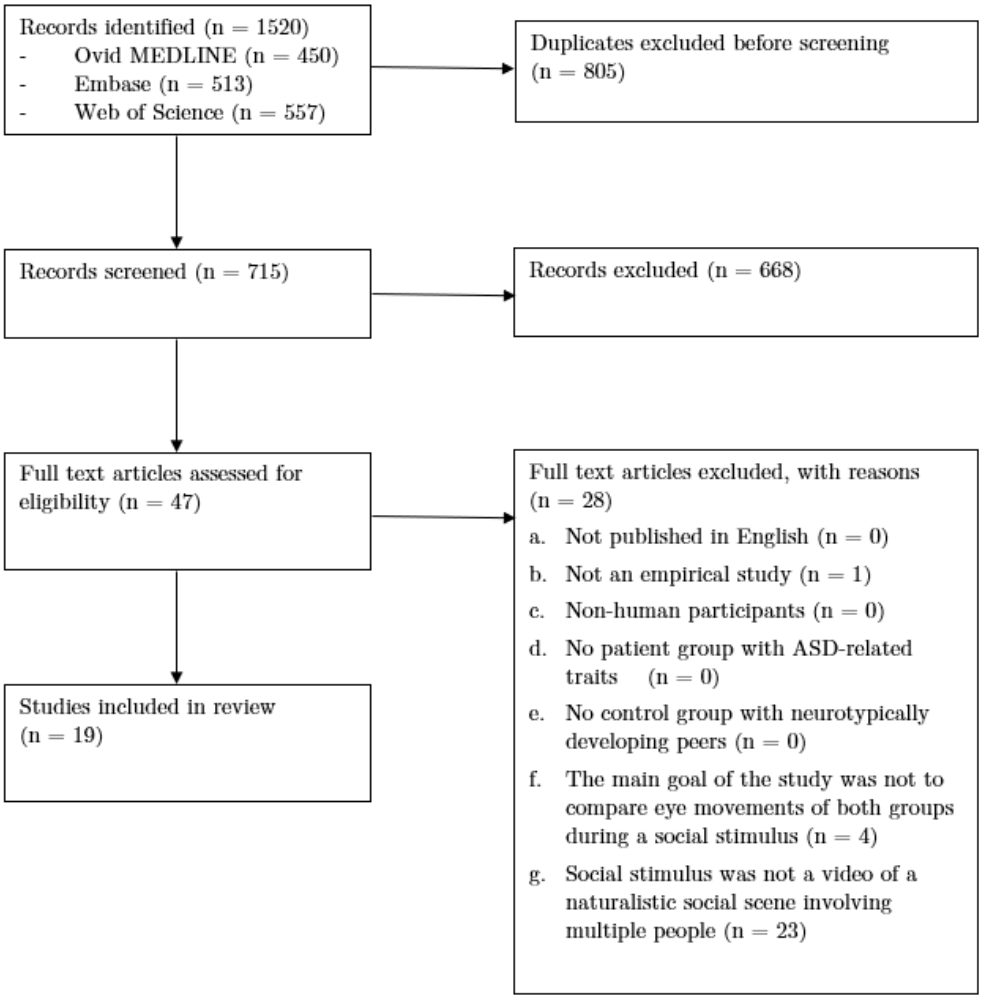


Figure 1: PRISMA flow diagram of the search strategy.

in this review. The hands area of interest included the activity area when applicable.

AOI: Face

Eight studies found a lower fixation time on the face in ASD participants, [23–30] while four studies were unable to identify any difference in fixation time compared to TD participants. [31–34] One study observed a negative correlation with the ADOS severity score, indicating better social functioning when a participant fixates more on the face. [24] Three other studies did not notice a correlation with autism-related scales. [30, 32, 33] A larger effect size was observed when stimuli consisted of more socially complex scenes [26] and when a person in the video moved their head a lot [25].

AOI: Eyes

Three studies reported a lower fixation time on the eyes in ASD participants, [23, 27, 35] and one study could not identify a difference. [30] A positive correlation with MASC scores was noticed in one study, implying a relationship between better social functioning and eye fixation. [36] Three studies did not observe a correlation with autism-related scales. [27, 30, 35]

AOI: Mouth

The evidence for a difference in mouth fixation between individuals with ASD and TD is contradictory, with two articles showing a lower fixation time in ASD [23, 27] and one article showing a higher fixation time in ASD. [35] Furthermore, one study reported a positive correlation with social functioning scales, [35] and one reported no correlation with social functioning scales. [36] However, the rationale behind these conflicting results might be provided by Rice et al., as they found that IQ profiles moderated the correlation between fixation time on the mouth and social functioning. [27]

Table 1: Outcome parameters used in the literature. AOI = Area of Interest. ASD = Autism Spectrum Disorder. ADOS = Autism Diagnostic Observation Schedule. SRS = Social Responsiveness Scale. MASC = Multidimensional Anxiety Scale for Children. VABS-E = Vineland Adaptive Behaviour Scales Expanded edition.

AOI	Face	Lower fixation time in ASD [23–30] No difference in fixation time [31–34] Negative correlation with ADOS [24] No correlation with ADOS / SRS [30, 32, 33]
	Eyes	Lower fixation time in ASD [23, 27, 35] No difference in fixation time [30] Positive correlation with MASC [36] No correlation with ADOS / VABS-E [27, 30, 35]
	Mouth	Lower fixation time in ASD [23, 27] Higher fixation time in ASD [35] Positive correlation with social functioning [35] No correlation with MASC scores [36] Correlation with social functioning moderated by IQ profiles [27]
	Body	Higher fixation time in ASD [27, 35] No difference in fixation time [29, 34] No correlation with ADOS / VABS-E / MASC [27, 35, 36]
	Hands	Higher fixation time in ASD [29] No difference in fixation time [26, 33] No correlation with SRS [33]
	Distracting object	Higher fixation time in ASD [27, 34, 35] No difference in fixation time [26, 29, 31] Positive correlation with ADOS [27, 35] No correlation with VABS-E / MASC [35, 36]
	Background	No difference in fixation time [33] No correlation with SRS [33]
Gaze-speech correlation		Gaze was not kept at speaker till end of speech in ASD [23] Less predictive saccades in ASD [23, 25] Proportionally less fixation at speaker than non-speaker in ASD[25] Lower gaze-speech correlation in ASD[37]
Scanpath variance		Higher scanpath variance in ASD [23, 30, 38] Positive correlation with ADOS [30] No correlation with SRS [38]
Saccadic speed		No difference in saccadic speed [39] No correlation with SRS [39]
Total looking time		Significantly lower in ASD [39] No correlation with SRS [39]
Time per fixation		Significantly lower in ASD [25]
Frame-by-frame analysis		Less fixation on person receiving a non-verbal request in ASD[32, 40] More fixation on sandwich after it fell in ASD[41] No frame-by-frame differences [31]

AOI: Body

A higher fixation time on the body in the ASD group was reported by two studies, [27, 35] whereas two others could not detect a difference. [29, 34] No correlation between fixation time on the body and social functioning scales has been detected. [27, 35, 36]

AOI: Hands, activity area

One study reported a higher fixation time on the hands in the ASD group, [29] whereas two other studies did not observe this difference. [26, 33] No correlation with the SRS could be detected. [33]

AOI: Distracting object

A higher fixation time on distracting objects in participants with ASD was reported in three studies. [27, 34, 35] Three studies did not observe this difference. [26, 29, 31] A positive correlation between fixation time on distracting objects and the ADOS severity score was found by two studies, indicating worse social functioning in participants who fixated more on distracted objects. [27, 35] No correlation was observed between fixation time on distracting objects and scores on the VABS-E or MASC. [35, 36]

AOI: Background

No significant differences were observed between the ASD and TD group in background fixation time. [33] Additionally, no correlation with the SRS was detected. [33]

Gaze-speech correlation

Several pieces of evidence point towards a difference in gaze-speech correlation between individuals with ASD and TD. Individuals with ASD are less likely to keep their gaze at the speaker till the end of their speech. [23] In line with this finding it was found that people with ASD fixate proportionally less on the speaker than the non-speaker. [25] Furthermore, it has been reported by two studies that people with ASD exhibit less predivtive saccades than their TD peers. [23, 25] Lastly, Chang et al. found a discrepancy in gaze-speech correlation between ASD an TD. [37]

Scanpath variance

Three studies reported a higher scanpath variance in individuals with ASD. [23, 30, 38] In agreement with this, Avni et al. observed a correlation between scanpath variance and ADOS scores. [30] A correlation with the SRS could not be detected by Ramot et al. [38]

Saccadic speed

No difference in saccadic speed was detected between the ASD and TD group ($p = 0.08$). [39] No correlation between saccadic speed and the SRS has been detected by Shaffer et al. [39]

Total looking time

The total looking time, meaning the fixation time within screen boundaries, was lower in the ASD group compared to the TD group. [39] However, no correlation between total looking time and the SRS could be established. [39]

Time per fixation

The time per fixation was lower in the ASD group compared to the TD group. [25]

Frame-by-frame analysis

Four studies performed a frame-by-frame analysis, exploring in which part of the provided social scene gaze differences could be observed. Two of those studies reported that individuals with ASD fixated less on a person receiving a non-verbal request. [32, 40] One study observed more fixation on a sandwich after it fell. [41] Kaliukhovich et al. could not find any differences in their frame-by-frame analysis. [29]

4 Discussion

The results indicate that individuals with ASD exhibit distinct eye movement patterns when viewing naturalistic social stimuli. Specifically, they tend to fixate less on socially informative areas such as the face and eyes while exhibiting increased attention toward less socially informative elements such as the body, hands, and distracting objects. The observed reduction in fixation time on the eyes and face is consistent with theories of social disengagement in ASD. The amygdala and associated neural pathways, which are crucial for processing facial expressions and emotional cues, have been shown to function atypically in individuals with ASD. [14] Furthermore, increased fixation on non-social elements is consistent with the preference for predictable, rule-based information processing in individuals with ASD. [42]

Intriguingly, the fixation time on the mouth appears to be moderated by IQ profiles. [27] In individuals with a verbal IQ that exceeds their non-verbal IQ, mouth fixation had a negative correlation with ADOS severity scores. This finding indicates that in this specific group, a greater tendency to fixate on the mouth is associated with better social functioning. Conversely, in the group with a low and evenly distributed IQ, no such correlation was observed. Notably, in the group with a high and evenly distributed IQ, a positive correlation was identified between mouth fixation and ADOS severity scores, suggesting that increased mouth fixation is associated with worse social functioning in this

subgroup. These findings underscore the complexity of social functioning and highlight the potential moderating role of IQ in this context.

Beyond the areas of interest, several studies examined non-AOI outcome measures, which provided additional insights into the eye movement behaviour of individuals with ASD. These measures included gaze-speech correlation, scanpath variance, saccadic speed, total looking time, time per fixation, and frame-by-frame analysis. Individuals with ASD demonstrated lower gaze-speech correlation, they were less likely to maintain gaze on a speaker until the conclusion of their speech and showed fewer predictive saccades, indicating difficulties in anticipating social cues. Higher scanpath variance, also termed idiosyncracy, in ASD participants suggests a more dispersed and less structured visual search pattern. This finding aligns with previous research showing that individuals with ASD exhibit more variable cortical responses as well as idiosyncratic distortion of the functional connectivity pattern. [43, 44] This link highlights how neurological atypicalities in ASD may manifest in observable behavior, though this remains a tentative interpretation requiring further investigation. While the saccadic speed outcome parameter did not reach statistical significance ($p = 0.08$), the p-value suggests a trend that could be worth exploring in future research. Additionally, a consistent reduction in total looking time was observed in individuals with ASD, suggesting a reduced engagement with the visual content of social scenes. Furthermore, frame-by-frame analysis revealed reduced fixation on individuals receiving non-verbal requests. The absence of gaze shifts toward a person receiving a non-verbal request might share its underlying mechanism with the lack of predictive saccades, as such gaze shifts are essentially predictive saccades. These findings underscore that a biomarker model should include include both AOI and non-AOI parameters to fully capture the complexities of gaze behaviour in ASD.

The small median sample size of 63 participants represents a significant limitation of the included studies. Furthermore, a methodological issue that was identified in the reviewed studies is the variation in the measurement of fixation time. Some studies report actual fixation time, defined as the absolute duration spent looking at a specific area, while others use proportional fixation time, defined as fixation time relative to the total screen fixation time. This distinction is critical because individuals with ASD tend to look at the screen for a significantly shorter duration overall than their TD counterparts. [39] Using proportional fixation time may, therefore, overestimate their engagement with social stimuli. It would be recommended for future research to use the actual fixation time, rather than the proportional fixation time.

In order to ascertain whether eye tracking is beneficial solely for diagnostic purposes or also for monitoring the effect of interventions, future studies should focus on whether gaze behaviour undergoes alterations over time in response to interventions. Longitudinal research could assist in identifying whether gaze patterns can be modified through targeted therapies and whether such changes are associated with improved social functioning. Another important avenue for research is exploring cultural and linguistic influences on gaze behaviour. Given that social communication norms vary across cultures, it is possible that gaze preferences in ASD are influenced by sociocultural factors. Studies with more diverse participant populations could help elucidate these effects and inform the adaptation of stimuli and variables for varied cultural contexts. Moreover, advanced machine learning models may enable individualized assessments, which could inform personalized interventions. Finally, from a clinical implementation perspective, integrating eye tracking into ASD screening procedures requires addressing considerations such as cost, scalability, and the clinical benefits in comparison to alternative diagnostic tools. While eye tracking technology has become more affordable in recent years, [45] further research is necessary to determine the optimal point in the patient journey at which this tool provides the most benefit.

In conclusion, individuals with ASD have been shown to exhibit reduced fixation on social cues, increased attention to non-social elements, and altered gaze patterns. These patterns may be related to atypical amygdala function, idiosyncratic cortical responses and idiosyncratic distortion of the functional connectivity pattern. Moving forward, it is crucial to further investigate the relevance of combinations of these outcome parameters within larger sample sizes and to examine the effects of various interventions.

References

1. Zeidan, J. *et al.* Global prevalence of autism: A systematic review update. en. *Autism Res.* **15,** 778–790 (May 2022).
2. Matuskey, D. *et al.* 11C-UCB-J PET imaging is consistent with lower synaptic density in autistic adults. en. *Mol. Psychiatry* (Oct. 2024).
3. Gandal, M. J. *et al.* Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. en. *Nature* **611,** 532–539 (Nov. 2022).
4. Nadeem, M. S. *et al.* Autism - A comprehensive array of prominent signs and symptoms. en. *Curr. Pharm. Des.* **27,** 1418–1433 (2021).
5. Park, S. H. *et al.* Disability, functioning, and quality of life among treatment-seeking young autistic adults and its relation to depression, anxiety, and stress. en. *Autism* **23,** 1675–1686 (Oct. 2019).
6. Movsessian, T. & Osoba, T. A. Association between therapeutic interventions and quality of life in people with autism. *J. Soc. Behav. Health Sci.* **16** (Nov. 2022).
7. Tedla, J. S. *et al.* Assessing the quality of life in children with autism spectrum disorder: a cross-sectional study of contributing factors. en. *Front. Psychiatry* **15,** 1507856 (Dec. 2024).
8. Vivanti, G., Prior, M., Williams, K. & Dissanayake, C. Predictors of outcomes in autism early intervention: Why donâ€™t we know more? en. *Front. Pediatr.* **2,** 58 (June 2014).
9. Okoye, C. *et al.* Early diagnosis of autism spectrum disorder: A review and analysis of the risks and benefits. en. *Cureus* **15,** e43226 (Aug. 2023).
10. Bishop, S. L. & Lord, C. Commentary: Best practices and processes for assessment of autism spectrum disorder - the intended role of standardized diagnostic instruments. en. *J. Child Psychol. Psychiatry* **64,** 834–838 (May 2023).
11. Huda, E. *et al.* Screening tools for autism in culturally and linguistically diverse paediatric populations: a systematic review. en. *BMC Pediatr.* **24,** 610 (Sept. 2024).
12. Lefort-Besnard, J. *et al.* Patterns of autism symptoms: hidden structure in the ADOS and ADI-R instruments. en. *Transl. Psychiatry* **10,** 257 (July 2020).
13. Hus, Y. & Segal, O. Challenges surrounding the diagnosis of autism in children. en. *Neuropsychiatr. Dis. Treat.* **17,** 3509–3529 (Dec. 2021).
14. Papagiannopoulou, E. A., Chitty, K. M., Hermens, D. F., Hickie, I. B. & Lagopoulos, J. A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. en. *Soc. Neurosci.* **9,** 610–632 (July 2014).
15. Frazier, T. W. *et al.* Evidence-based use of scalable biomarkers to increase diagnostic efficiency and decrease the lifetime costs of autism. en. *Autism Res.* **14,** 1271–1283 (June 2021).
16. Bölte, S. *et al.* How can clinicians detect and treat autism early? Methodological trends of technology use in research. en. *Acta Paediatr.* **105,** 137–144 (Feb. 2016).
17. Klin, A. Biomarkers in autism spectrum disorder: Challenges, advances, and the need for biomarkers of relevance to public health. en. *Focus (Am. Psychiatr. Publ.)* **16,** 135–142 (Apr. 2018).
18. Speer, L. L., Cook, A. E., McMahon, W. M. & Clark, E. Face processing in children with autism: effects of stimulus contents and type. en. *Autism* **11,** 265–277 (May 2007).
19. Chita-Tegmark, M. Social attention in ASD: A review and meta-analysis of eye-tracking studies. en. *Res. Dev. Disabil.* **48,** 79–93 (Jan. 2016).
20. Grossman, R. B., Zane, E., Mertens, J. & Mitchell, T. Facetime vs. Screentime: Gaze patterns to live and video social stimuli in adolescents with ASD. en. *Sci. Rep.* **9,** 12643 (Sept. 2019).
21. Robain, F. *et al.* Measuring social orienting in preschoolers with autism spectrum disorder using cartoons stimuli. en. *J. Psychiatr. Res.* **156,** 398–405 (Dec. 2022).
22. Parker, T. C. *et al.* Neural and visual processing of social gaze cueing in typical and ASD adults. en. *medRxiv* (Feb. 2023).
23. Nakano, T. *et al.* Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour. en. *Proc. Biol. Sci.* **277,** 2935–2943 (Oct. 2010).

24. Zantinge, G., van Rijn, S., Stockmann, L. & Swaab, H. Psychophysiological responses to emotions of others in young children with autism spectrum disorders: Correlates of social functioning. *Autism Res.* **10,** 1499–1509 (Sept. 2017).
25. Von Hofsten, C., Uhlig, H., Adell, M. & Kochukhova, O. How children with autism look at events. en. *Res. Autism Spectr. Disord.* **3,** 556–569 (Apr. 2009).
26. Parish-Morris, J. *et al.* Adaptation to different communicative contexts: an eye tracking study of autistic adults. en. *J. Neurodev. Disord.* **11,** 5 (Apr. 2019).
27. Rice, K., Moriuchi, J. M., Jones, W. & Klin, A. Parsing heterogeneity in autism spectrum disorders: visual scanning of dynamic social scenes in school-aged children. en. *J. Am. Acad. Child Adolesc. Psychiatry* **51,** 238–248 (Mar. 2012).
28. Constantino, J. N. *et al.* Infant viewing of social scenes is under genetic control and is atypical in autism. en. *Nature* **547,** 340–344 (July 2017).
29. Kaliukhovich, D. A. *et al.* Social attention to activities in children and adults with autism spectrum disorder: effects of context and age. en. *Mol. Autism* **11,** 79 (Oct. 2020).
30. Avni, I. *et al.* Children with autism observe social interactions in an idiosyncratic manner. en. *Autism Res.* **13,** 935–946 (June 2020).
31. Yoshida, A. *et al.* Eye gaze and cerebral blood flow activation while watching social movies in children with autism spectrum disorder. en. *J. Brain Sci.* **51,** 47–76 (Feb. 2022).
32. Viktorsson, C., Bölte, S. & Falck-Ytter, T. How 18-month-olds with later autism look at other children interacting: The timing of gaze allocation. en. *J. Autism Dev. Disord.* **54,** 4091–4101 (Nov. 2024).
33. Greene, R. K. *et al.* Dynamic eye tracking as a predictor and outcome measure of social skills intervention in adolescents and adults with autism spectrum disorder. en. *J. Autism Dev. Disord.* **51,** 1173–1187 (Apr. 2021).
34. Tang, J. S. Y., Chen, N. T. M., Falkmer, M., Blte, S. & Girdler, S. Atypical visual processing but comparable levels of emotion recognition in adults with autism during the processing of social scenes. en. *J. Autism Dev. Disord.* **49,** 4009–4018 (Oct. 2019).
35. Klin, A., Jones, W., Schultz, R., Volkmar, F. & Cohen, D. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. en. *Arch. Gen. Psychiatry* **59,** 809–816 (Sept. 2002).
36. Müller, N., Baumeister, S., Dziobek, I., Banaschewski, T. & Poustka, L. Validation of the Movie for the Assessment of Social Cognition in adolescents with ASD: Fixation duration and pupil dilation as predictors of performance. en. *J. Autism Dev. Disord.* **46,** 2831–2844 (Sept. 2016).
37. Chang, Z. *et al.* Computational methods to measure patterns of gaze in toddlers with autism spectrum disorder. en. *JAMA Pediatr.* **175,** 827–836 (Aug. 2021).
38. Ramot, M., Walsh, C., Reimann, G. E. & Martin, A. Distinct neural mechanisms of social orienting and mentalizing revealed by independent measures of neural and eye movement typicality. en. *Commun. Biol.* **3,** 48 (Jan. 2020).
39. Shaffer, R. C. *et al.* Brief report: Diminished gaze preference for dynamic social interaction scenes in youth with autism spectrum disorders. en. *J. Autism Dev. Disord.* **47,** 506–513 (Feb. 2017).
40. Falck-Ytter, T., von Hofsten, C., Gillberg, C. & Fernell, E. Visualization and analysis of eye movement data from children with typical and atypical development. en. *J. Autism Dev. Disord.* **43,** 2249–2258 (Oct. 2013).
41. Lönnqvist, L. *et al.* How young adults with autism spectrum disorder watch and interpret pragmatically complex scenes. en. *Q. J. Exp. Psychol. (Hove)* **70,** 2331–2346 (Nov. 2017).
42. Brosnan, M. & Ashwin, C. Thinking, fast and slow on the autism spectrum. en. *Autism* **27,** 1245–1255 (July 2023).
43. Byrge, L., Dubois, J., Tyszka, J. M., Adolphs, R. & Kennedy, D. P. Idiosyncratic brain activation patterns are associated with poor social comprehension in autism. en. *J. Neurosci.* **35,** 5837–5850 (Apr. 2015).
44. Hahamy, A., Behrmann, M. & Malach, R. The idiosyncratic brain: distortion of spontaneous connectivity patterns in autism spectrum disorder. en. *Nat. Neurosci.* **18,** 302–309 (Feb. 2015).
45. Guo, R., Kim, N. & Lee, J. Empirical insights into eye-tracking for design evaluation: Applications in visual communication and new media design. en. *Behav. Sci. (Basel)* **14** (Dec. 2024).

Appendix A: Search query

MEDLINE (exp * Autism Spectrum Disorder / OR ((Social ADJ3 Responsive*)).ab,ti,kw. OR (asd OR autis* OR Asperger* OR rett-sundrome* OR pdd-nos).ti.) AND (Eye Movements / OR Eye Movement Measurements / OR Eye-Tracking Technology / OR Fixation, Ocular / OR (((eye* OR ocular OR gaze*) ADJ3 (movement* OR tracking OR measurement* OR analys* OR pattern* OR cueing* OR cues OR estimation* OR fixation* OR follow*)) OR eyetrack*).ab,ti,kw. OR (gaze).ti.) AND (Video Recording/ OR (videorecor* OR video OR videos OR (live ADJ3 (interaction* OR observation* OR situation)) OR (social* ADJ3 (stimul* OR event* OR interaction* OR scene* OR task*)) OR (dynamic* ADJ3 setting*) OR ((viewing OR watching* OR complex*) ADJ3 scene*)).ab,ti,kw. OR (dynamic* OR stimulus*).ti.) NOT (* Attention Deficit Disorder with Hyperactivity / OR (attention-deficit-hyperactivity OR adhd).ti.) NOT (news OR congres* OR abstract* OR book* OR chapter* OR dissertation abstract*).pt. NOT (Systematic Review / OR Meta-Analysis / OR Case Reports / OR (systematic-review* OR meta-analy* OR case-report*).ti.) NOT (exp animals/ NOT humans/) AND english.la.

Embase ('Social Responsiveness Scale'/exp OR autism/mj/exp OR 'autism screening questionnaire'/mj OR 'autism assessment'/mj/exp OR 'autism screening'/mj OR ((Social NEAR/3 Responsive*)):ab,ti,kw OR (asd OR autis* OR Asperger* OR rett-sundrome* OR pdd-nos):ti) AND ('eye movement'/exp OR 'eye movement monitor'/exp OR 'eye tracking scan path'/de OR 'eye tracking technology'/de OR gaze/mj OR 'eye fixation'/de OR (((eye* OR ocular OR gaze*) NEAR/3 (movement* OR tracking OR measurement* OR analys* OR pattern* OR cueing* OR cues OR estimation* OR fixation* OR follow*)) OR eyetrack*):ab,ti,kw OR (gaze):ti) AND (videorecording/de OR 'social stimulation'/de OR (videorecor* OR video OR videos OR (live NEAR/3 (interaction* OR observation* OR situation)) OR (social* NEAR/3 (stimul* OR event* OR interaction* OR scene* OR task*)) OR (dynamic* NEAR/3 setting*) OR ((viewing OR watching* OR complex*) NEAR/3 scene*)):ab,ti,kw OR (dynamic* OR stimulus*):ti) NOT ('attention deficit hyperactivity disorder'/mj OR (attention-deficit-hyperactivity OR adhd):ti) NOT [conference abstract]/lim NOT ('systematic review'/de OR 'meta analysis'/de OR 'case report'/de OR (systematic-review* OR meta-analy* OR case-report*):ti) NOT ([animals]/lim NOT [humans]/lim) AND [english]/lim

Web of Science (TS=((Social NEAR/2 Responsive*)) OR TI=(asd OR autis* OR Asperger* OR rett-sundrome* OR pdd-nos)) AND (TS=(((eye* OR ocular OR gaze*) NEAR/2 (movement* OR tracking OR measurement* OR analys* OR pattern* OR cueing* OR cues OR estimation* OR fixation* OR follow*)) OR eyetrack*) OR TI=(gaze)) AND (TS=(videorecor* OR video OR videos OR (live NEAR/2 (interaction* OR observation* OR situation)) OR (social* NEAR/2 (stimul* OR event* OR interaction* OR scene* OR task*)) OR (dynamic* NEAR/2 setting*) OR ((viewing OR watching* OR complex*) NEAR/2 scene*)) OR TI=(dynamic* OR stimulus*)) NOT TI=((attention-deficit-hyperactivity OR adhd)) NOT TI=((systematic-review* OR meta-analy* OR case-report*)) NOT DT=(Meeting Abstract OR Meeting Summary) AND LA=(English)

B

Feature Distribution

Background Fixation

Table B.1: Log-likelihood and parameters for distribution fits on background fixation.

Participants	Distribution	LogLikelihood	μ	σ
All			0.02	0.06
	Normal	10726.84	0.02	0.00
	Beta	67431.86	0.01	0.00
	Gamma	**79444.00**	0.08	0.10
	Uniform	146.63	0.49	0.08

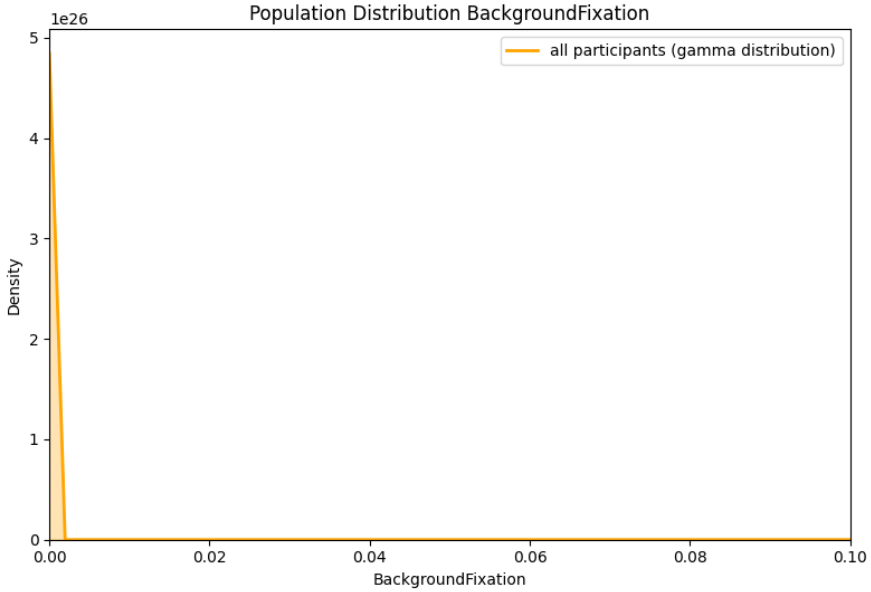


Figure B.1: Population distribution of background fixation.

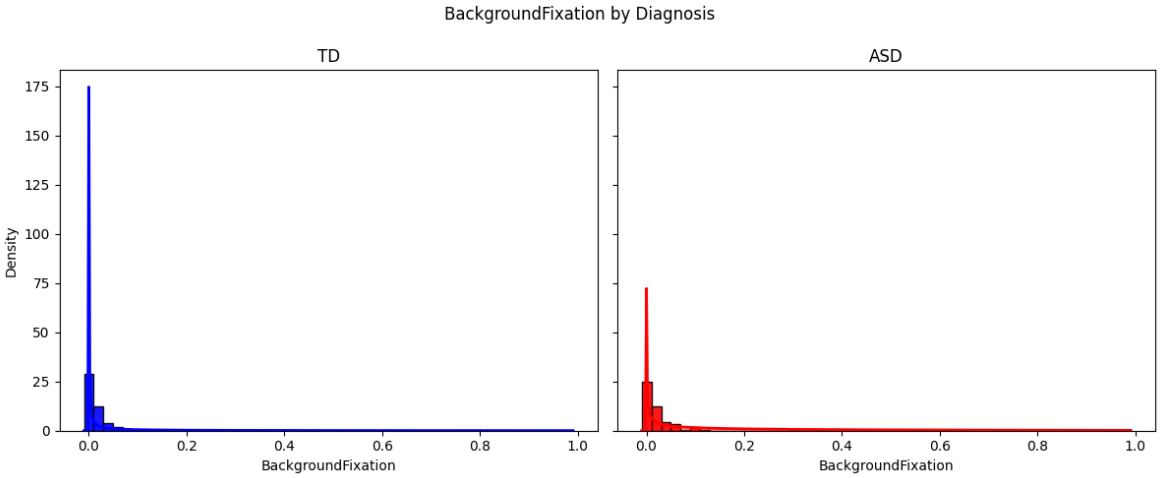


Figure B.2: Distribution by autism diagnosis, showing fitted distributions for each group.

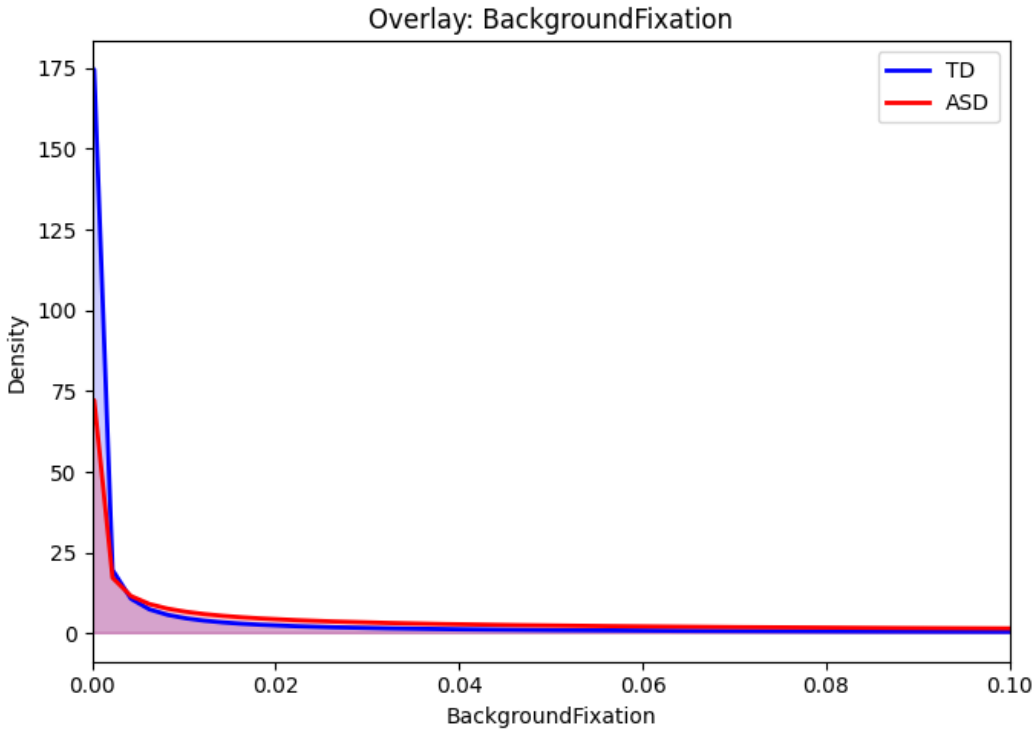


Figure B.3: Distribution overlays by autism diagnosis, showing fitted distributions for each group.

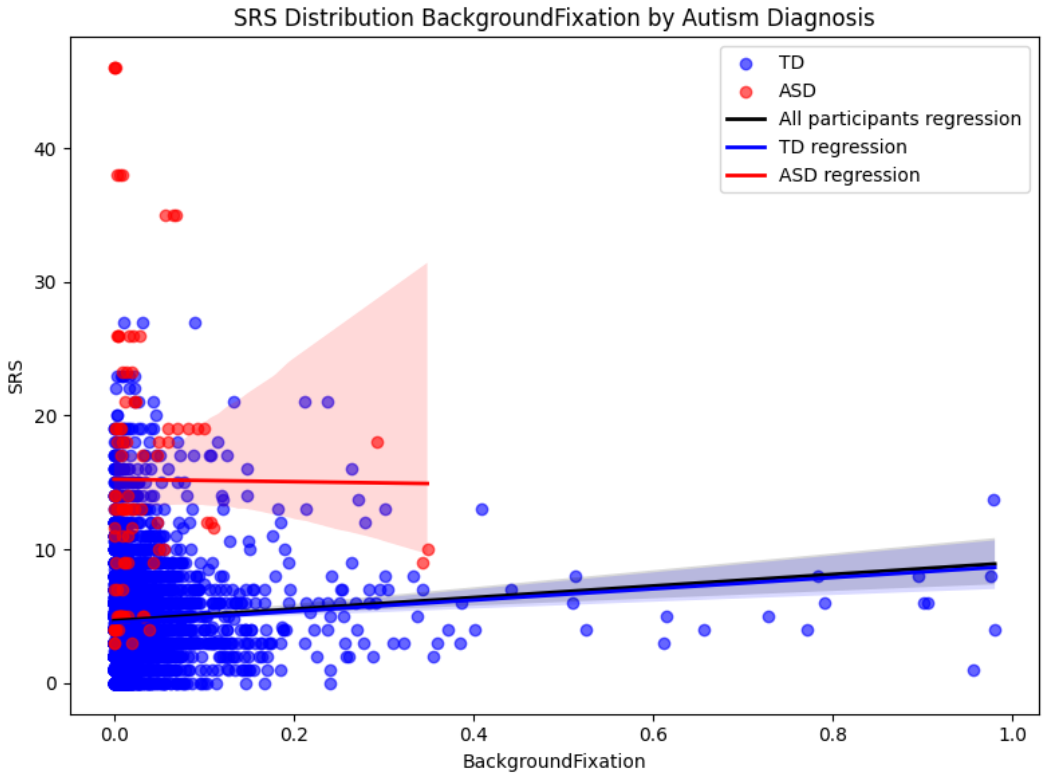


Figure B.4: Relationship between background fixation and SRS scores by diagnosis.

Body Fixation

Table B.2: Log-likelihood and parameters for distribution fits on body fixation.

Participants	Distribution	LogLikelihood	μ	σ
All			0.09	0.12
	Normal	5197.26	0.09	0.01
	Beta	9179.29	0.06	0.00
	Gamma	**17992.18**	0.09	0.01
	Uniform	378.46	0.48	0.08

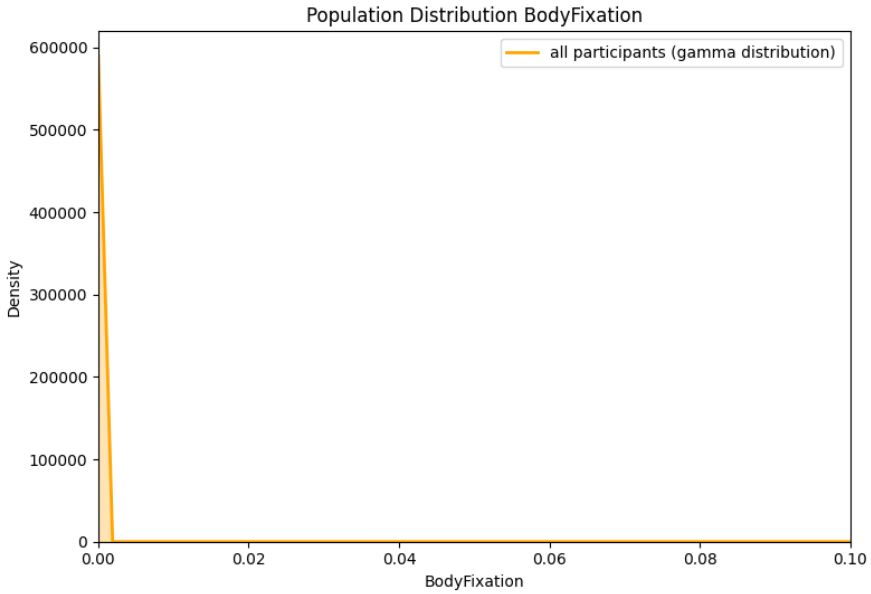


Figure B.5: Population distribution of body fixation.

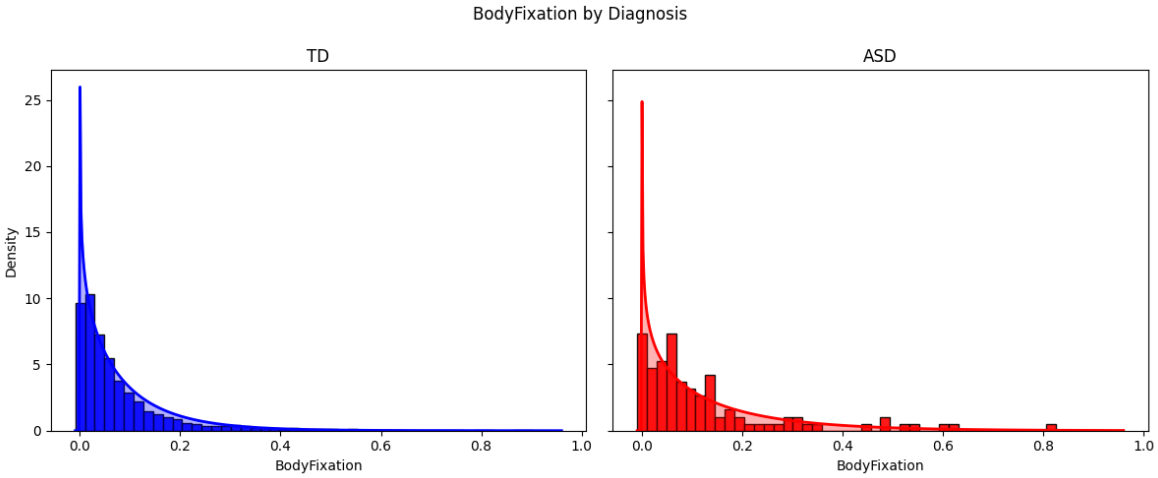


Figure B.6: Distribution by autism diagnosis, showing fitted distributions for each group.

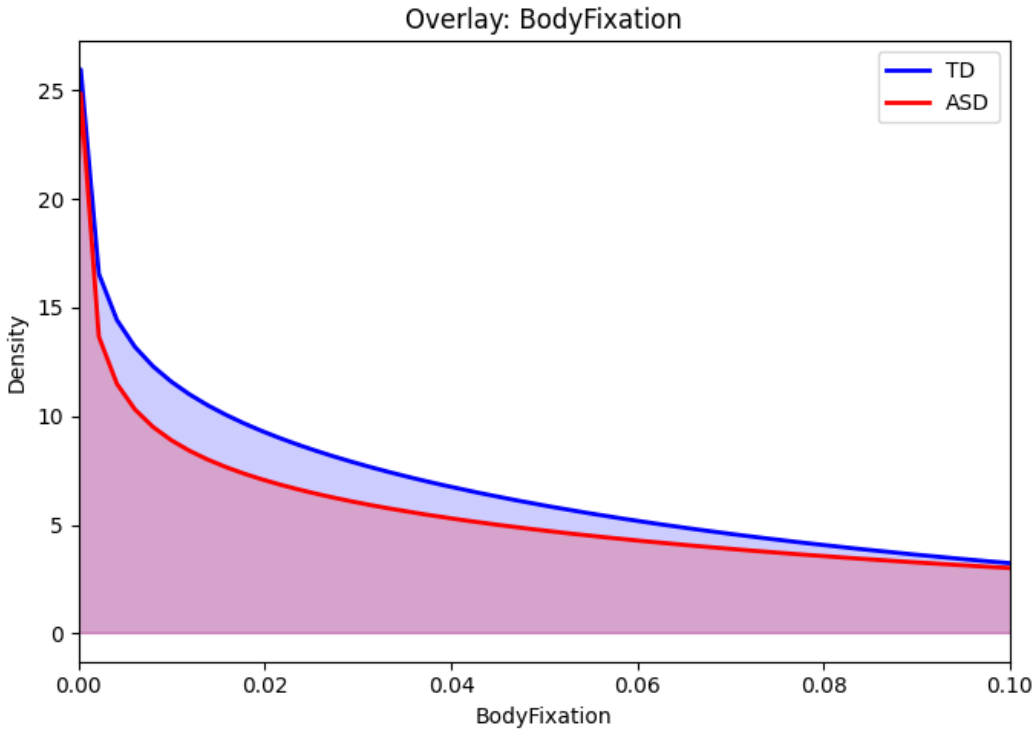


Figure B.7: Distribution overlays by autism diagnosis, showing fitted distributions for each group.

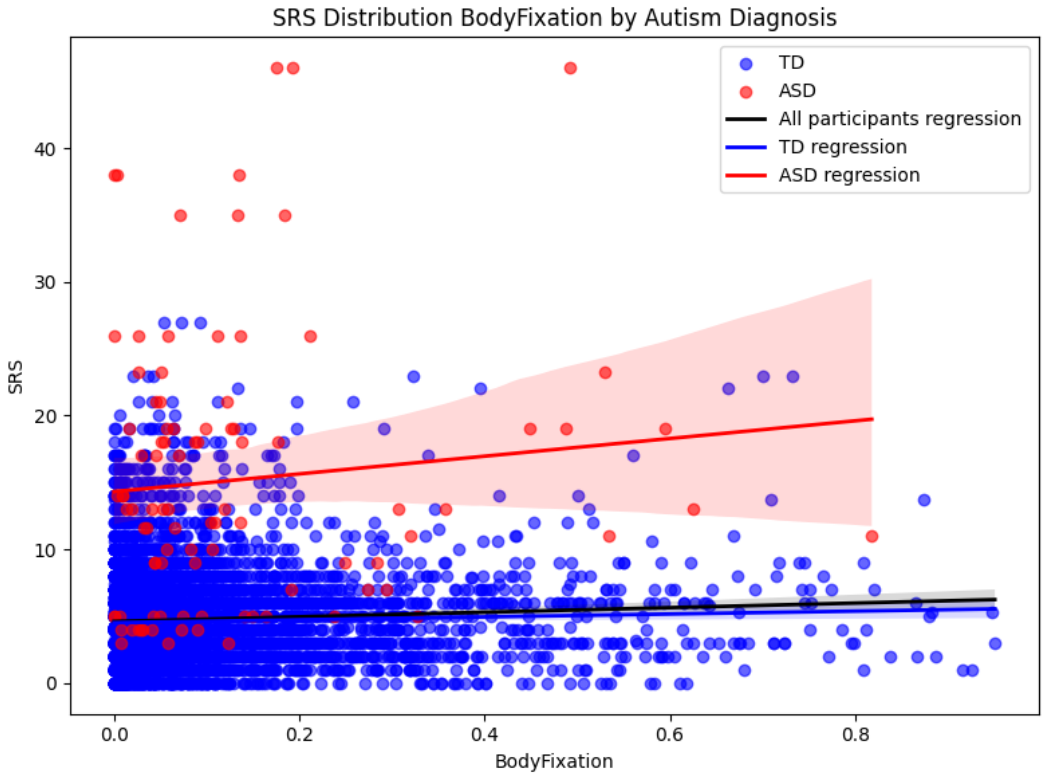


Figure B.8: Relationship between body fixation and SRS scores by diagnosis.

Eyes Fixation

Table B.3: Log-likelihood and parameters for distribution fits on eyes fixation.

Participants	Distribution	LogLikelihood	μ	σ
All			0.15	0.19
	Normal	1734.54	0.15	0.04
	Beta	33566.02	0.20	0.04
	Gamma	**37655.69**	0.08	0.01
	Uniform	452.57	0.47	0.07

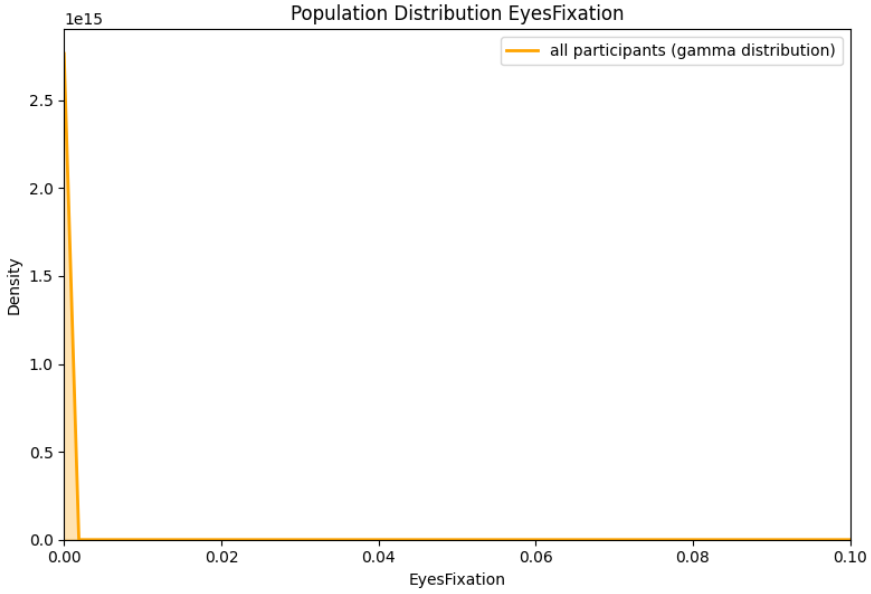


Figure B.9: Population distribution of eyes fixation.

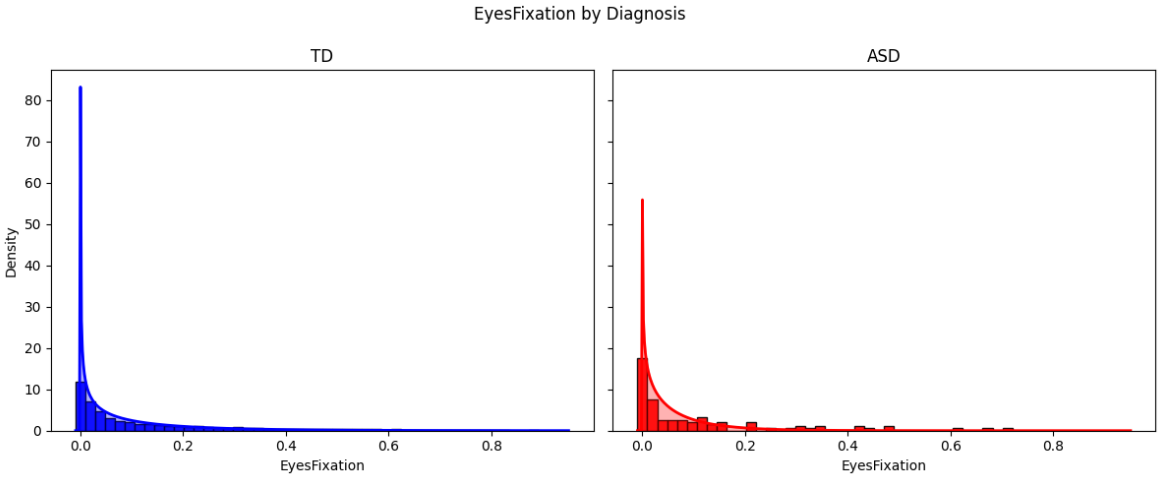


Figure B.10: Distribution by autism diagnosis, showing fitted distributions for each group.

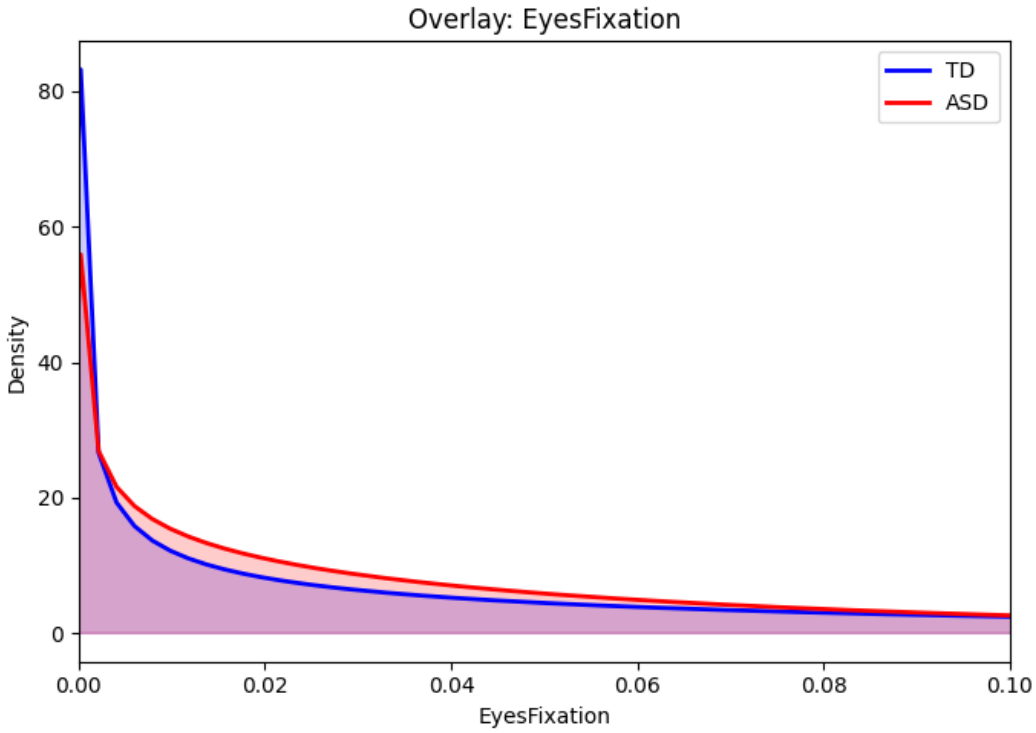


Figure B.11: Distribution overlays by autism diagnosis, showing fitted distributions for each group.

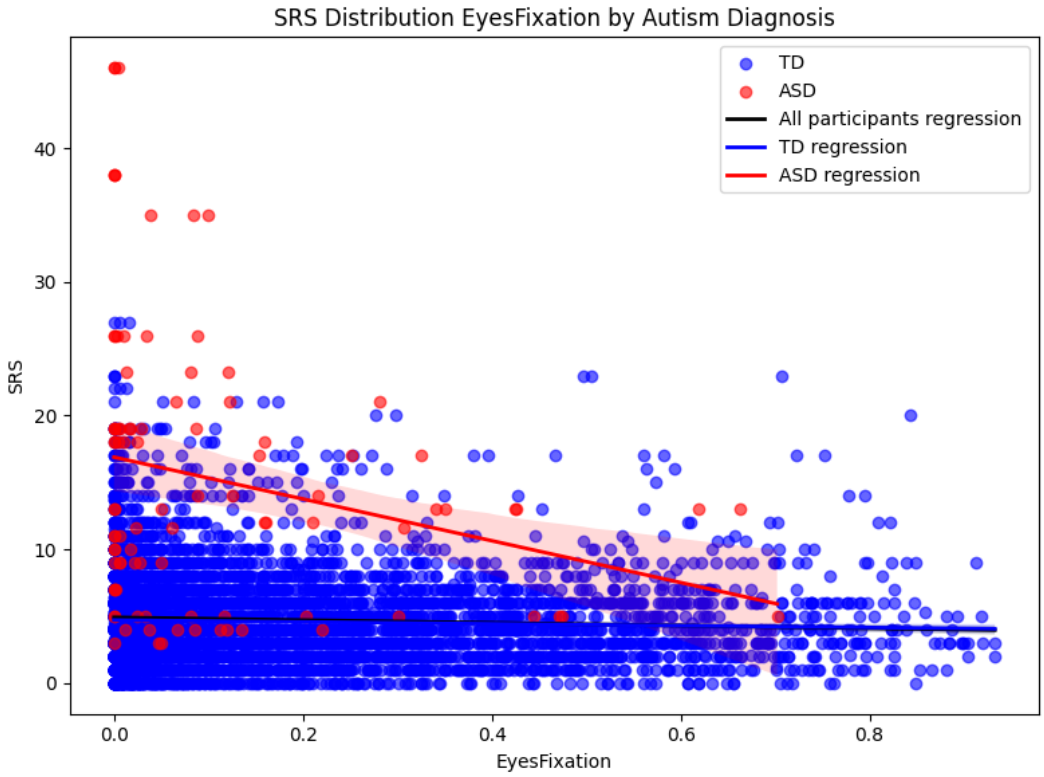


Figure B.12: Relationship between eyes fixation and SRS scores by diagnosis.

Face Fixation

Table B.4: Log-likelihood and parameters for distribution fits on face fixation.

Participants	Distribution	LogLikelihood	μ	σ
All			0.71	0.21
	Normal	1104.22	0.71	0.04
	Beta	**2212.62**	0.71	0.21
	Gamma	893.88	0.71	0.05
	Uniform	84.05	0.49	0.08

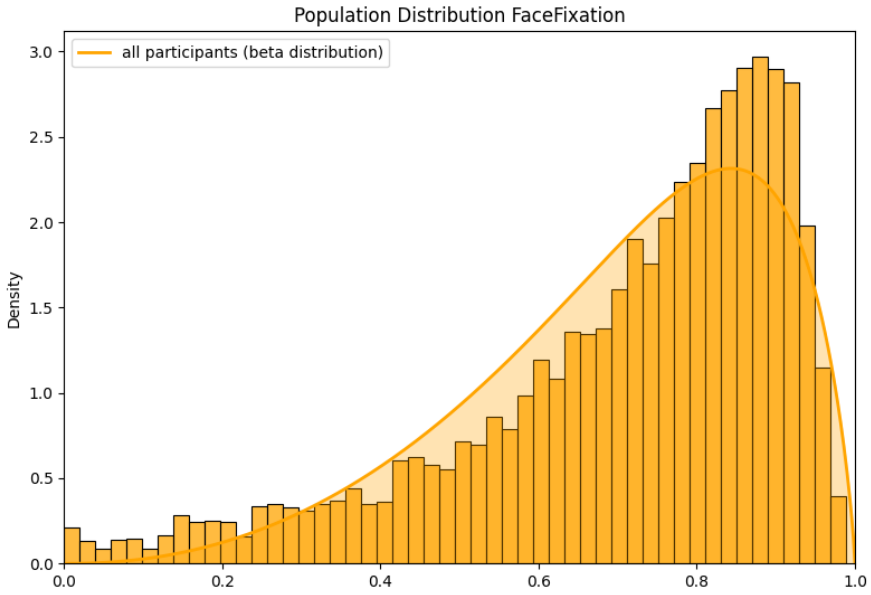


Figure B.13: Population distribution of face fixation.

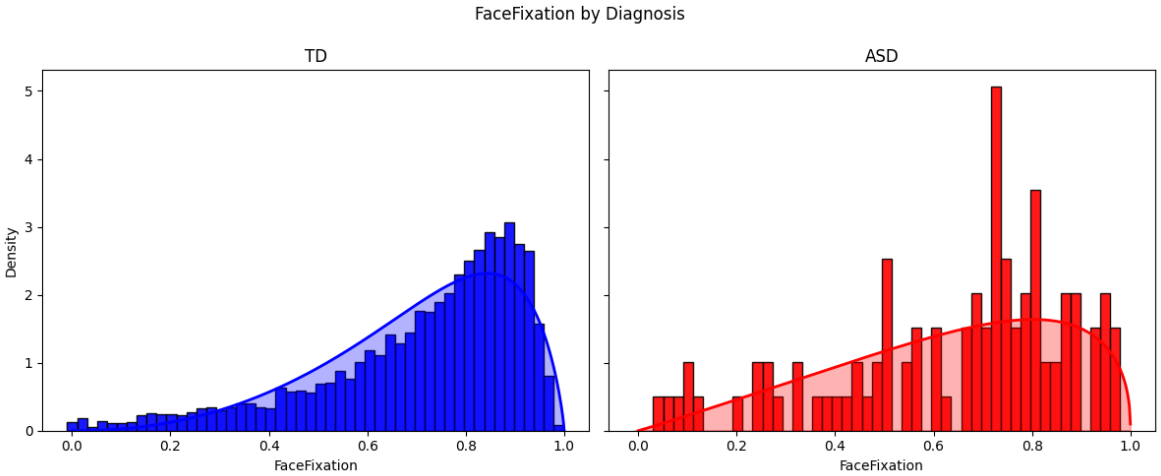


Figure B.14: Distribution by autism diagnosis, showing fitted distributions for each group.

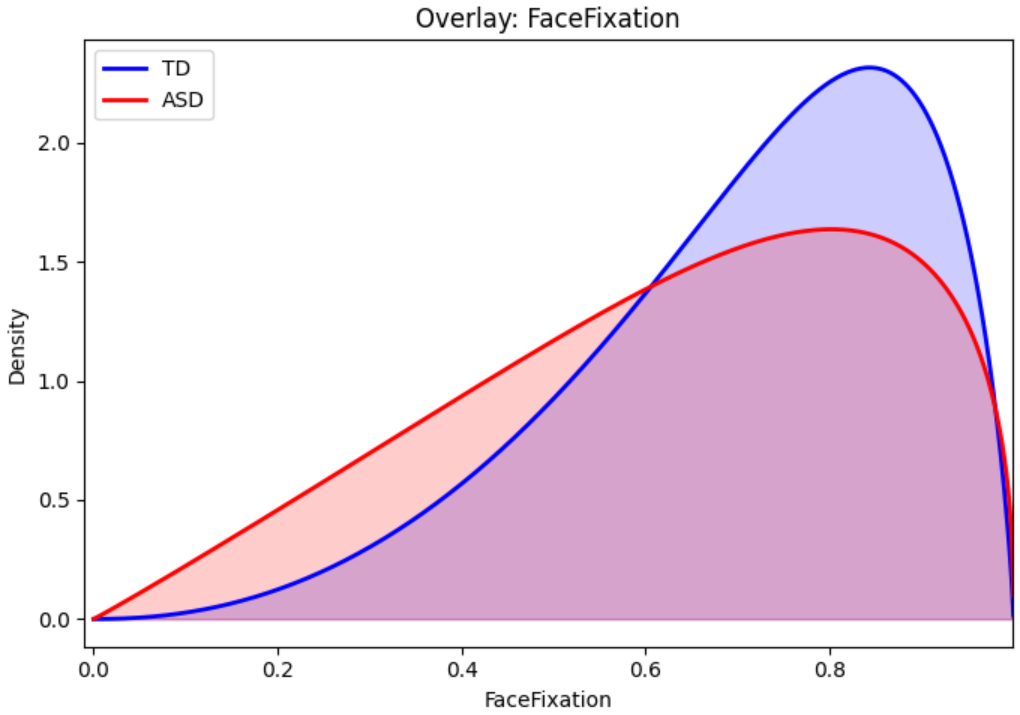


Figure B.15: Distribution overlays by autism diagnosis, showing fitted distributions for each group.

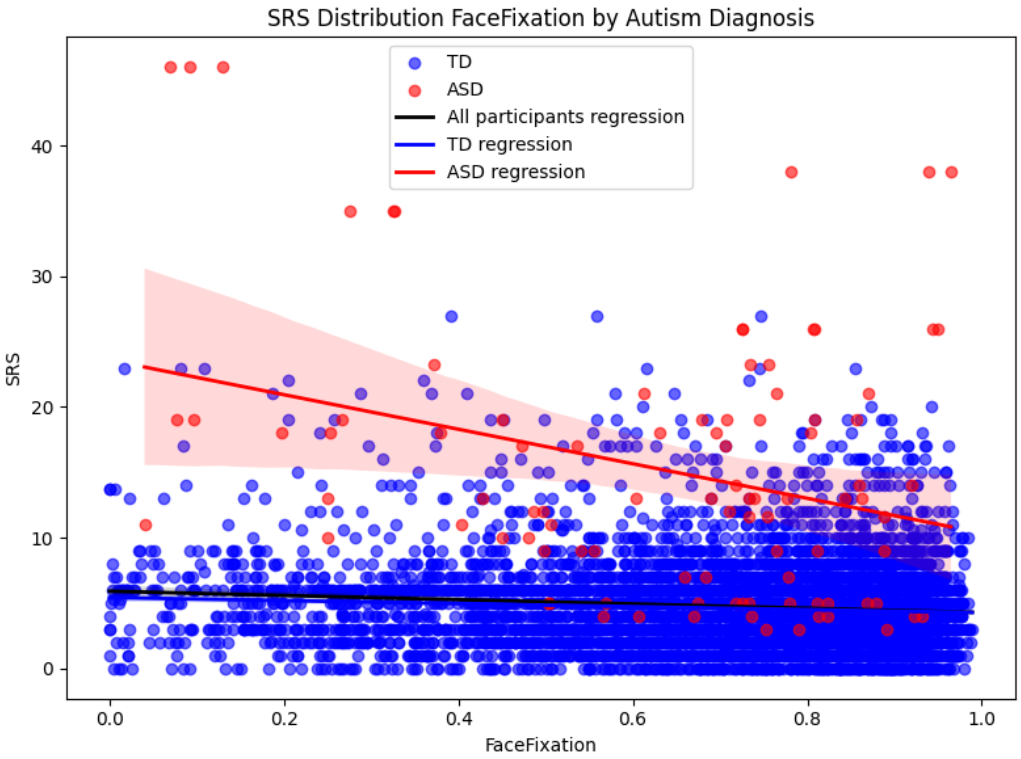


Figure B.16: Relationship between face fixation and SRS scores by diagnosis.

Mouth Fixation

Table B.5: Log-likelihood and parameters for distribution fits on mouth fixation.

Participants	Distribution	LogLikelihood	μ	σ
All			0.41	0.24
	Normal	-5.22	0.41	0.06
	Beta	**873.54**	0.39	0.06
	Gamma	-4.77	0.41	0.06
	Uniform	417.07	0.47	0.07

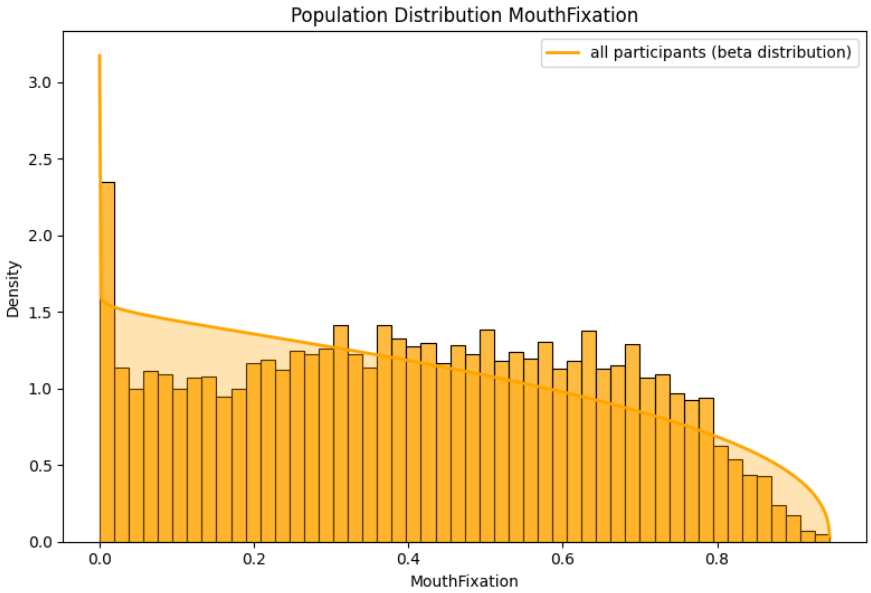


Figure B.17: Population distribution of mouth fixation.

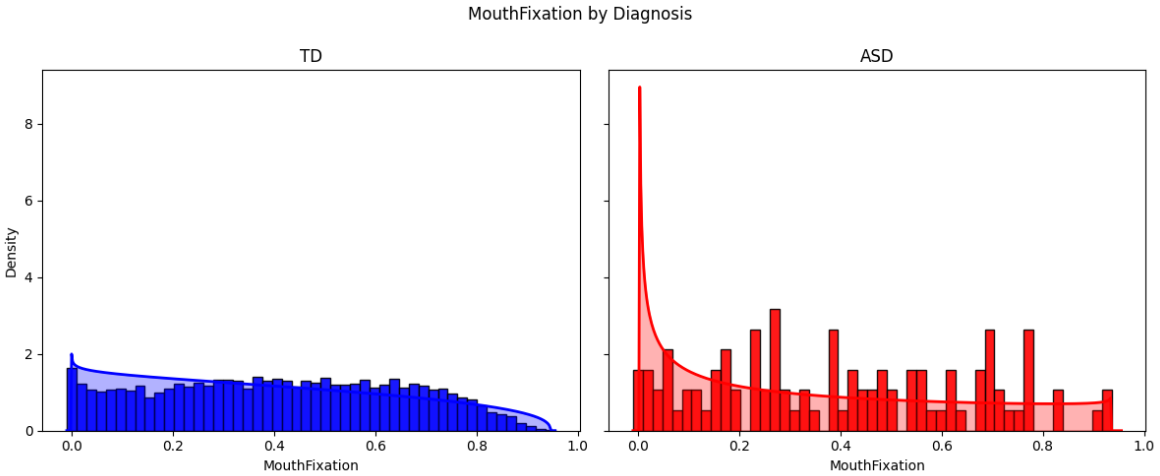


Figure B.18: Distribution by autism diagnosis, showing fitted distributions for each group.

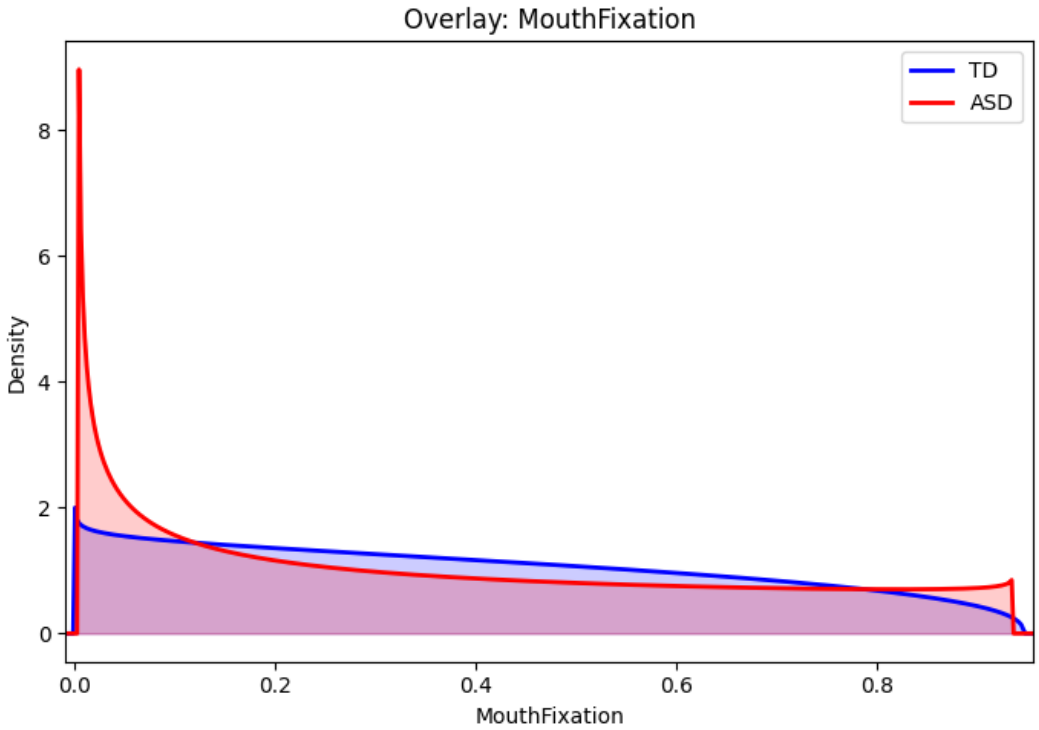


Figure B.19: Distribution overlays by autism diagnosis, showing fitted distributions for each group.

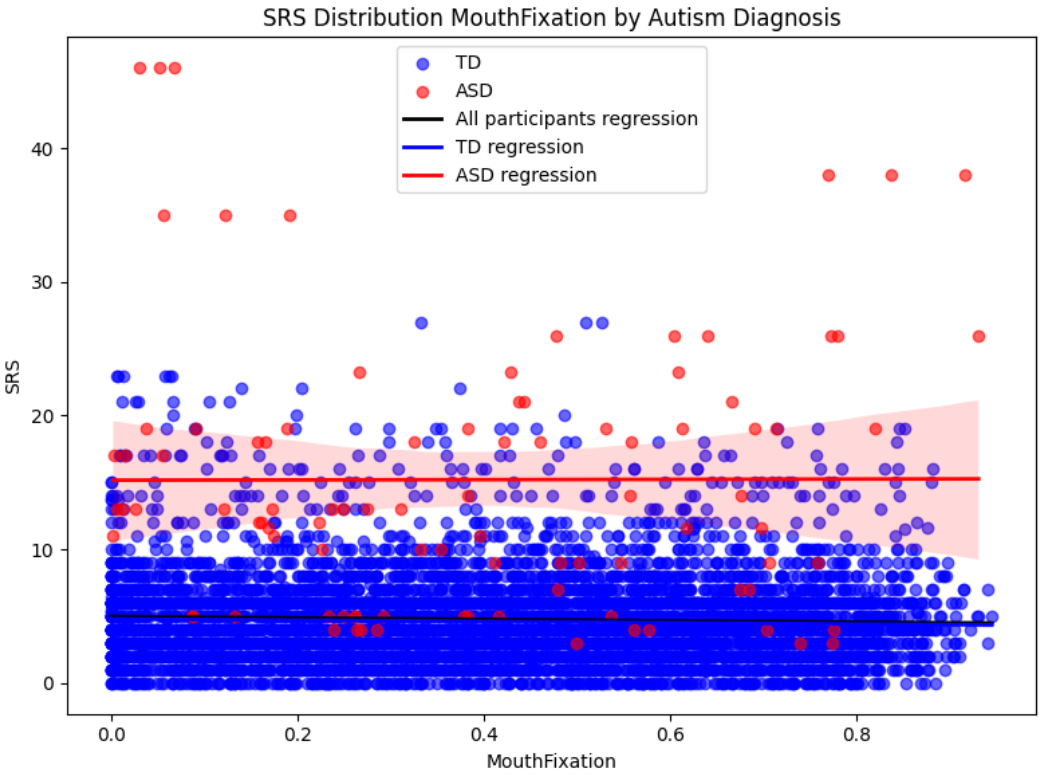


Figure B.20: Relationship between mouth fixation and SRS scores by diagnosis.

Non-Speaker Silhouette Correlation

Table B.6: Log-likelihood and parameters for distribution fits on non-speaker silhouette correlation.

Participants	Distribution	LogLikelihood	μ	σ
All			0.97	0.06
	Normal	10867.24	0.97	0.00
	Gamma	9102.61	0.97	0.01
	Beta	**36410.18**	0.97	0.00
	Uniform	2019.69	0.62	0.05

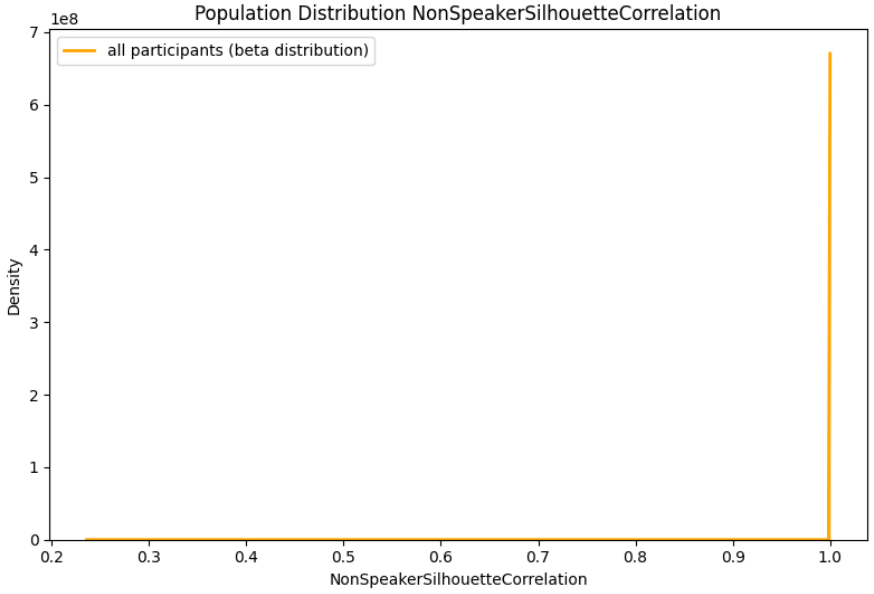


Figure B.21: Population distribution of non-speaker silhouette correlation.

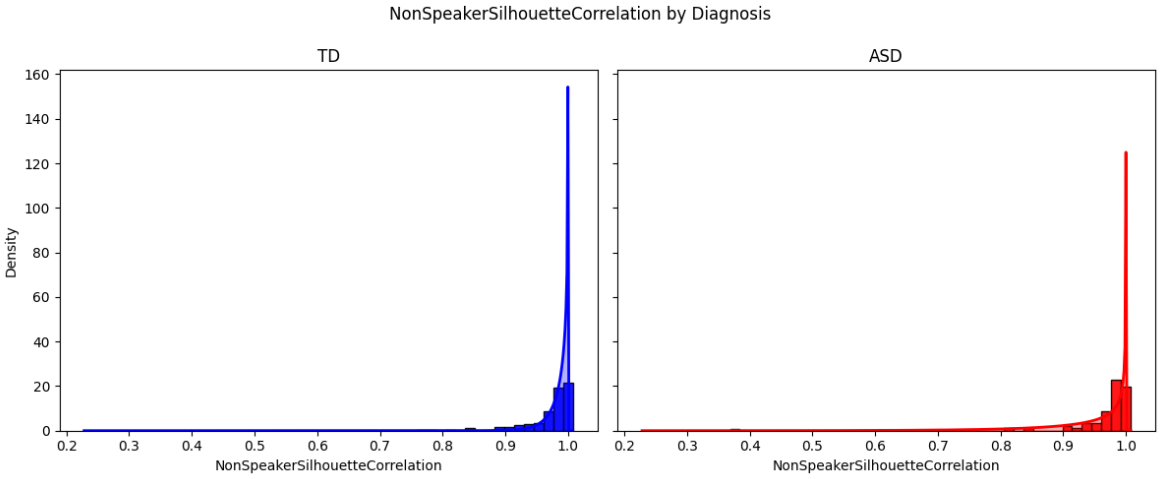


Figure B.22: Distribution by autism diagnosis, showing fitted distributions for each group.

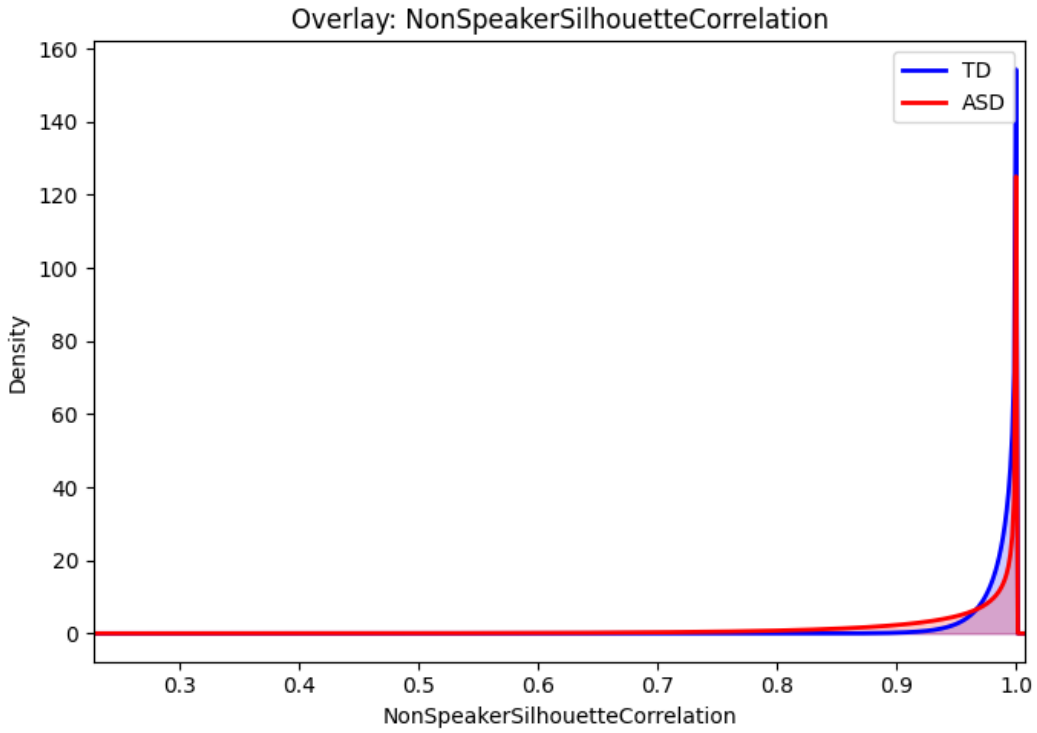


Figure B.23: Distribution overlays by autism diagnosis, showing fitted distributions for each group.

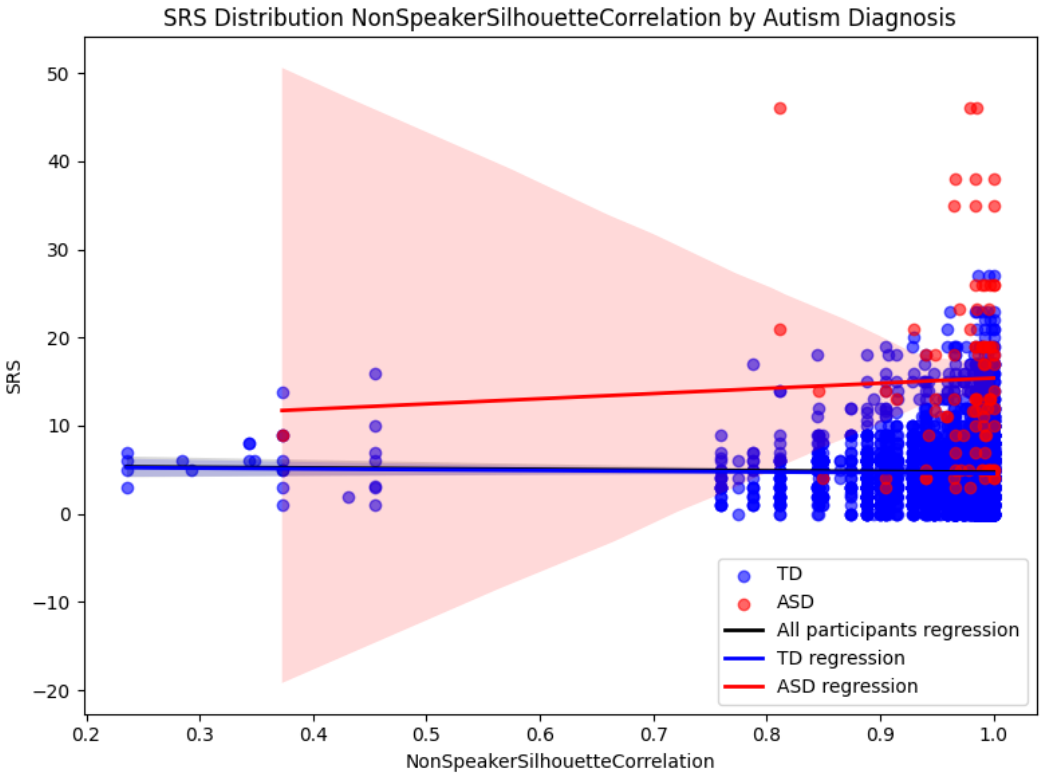


Figure B.24: Relationship between non-speaker silhouette correlation and SRS scores by diagnosis.

Non-Speaker Time Correlation

Table B.7: Log-likelihood and parameters for distribution fits on non-speaker time correlation.

Participants	Distribution	LogLikelihood	μ	σ
All			0.19	0.07
	Normal	9561.06	0.19	0.00
	Beta	9625.56	0.19	0.00
	Gamma	**9625.84**	0.19	0.00
	Uniform	4411.19	0.28	0.03

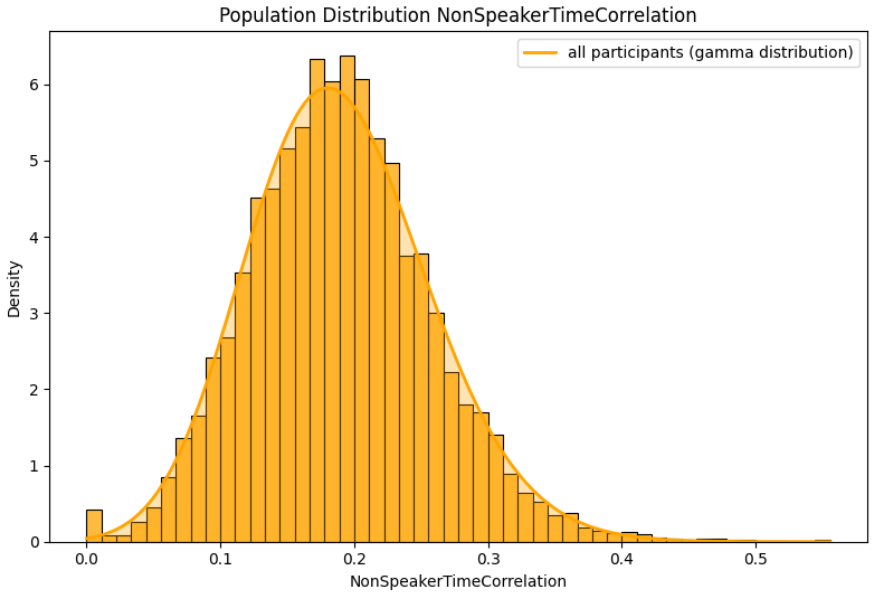


Figure B.25: Population distribution of non-speaker time correlation.

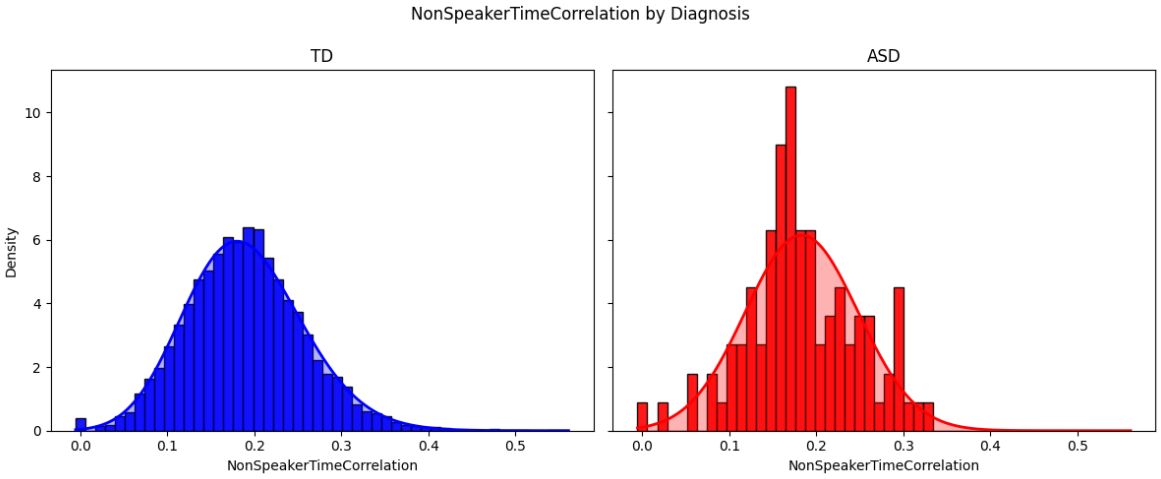


Figure B.26: Distribution by autism diagnosis, showing fitted distributions for each group.

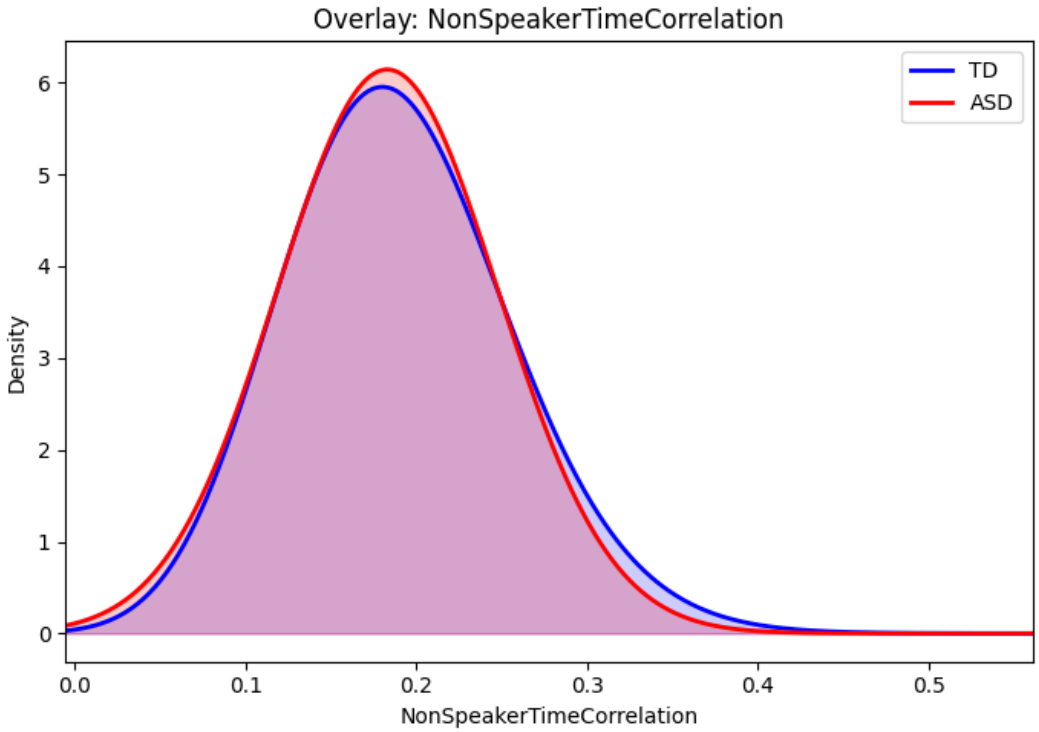


Figure B.27: Distribution overlays by autism diagnosis, showing fitted distributions for each group.

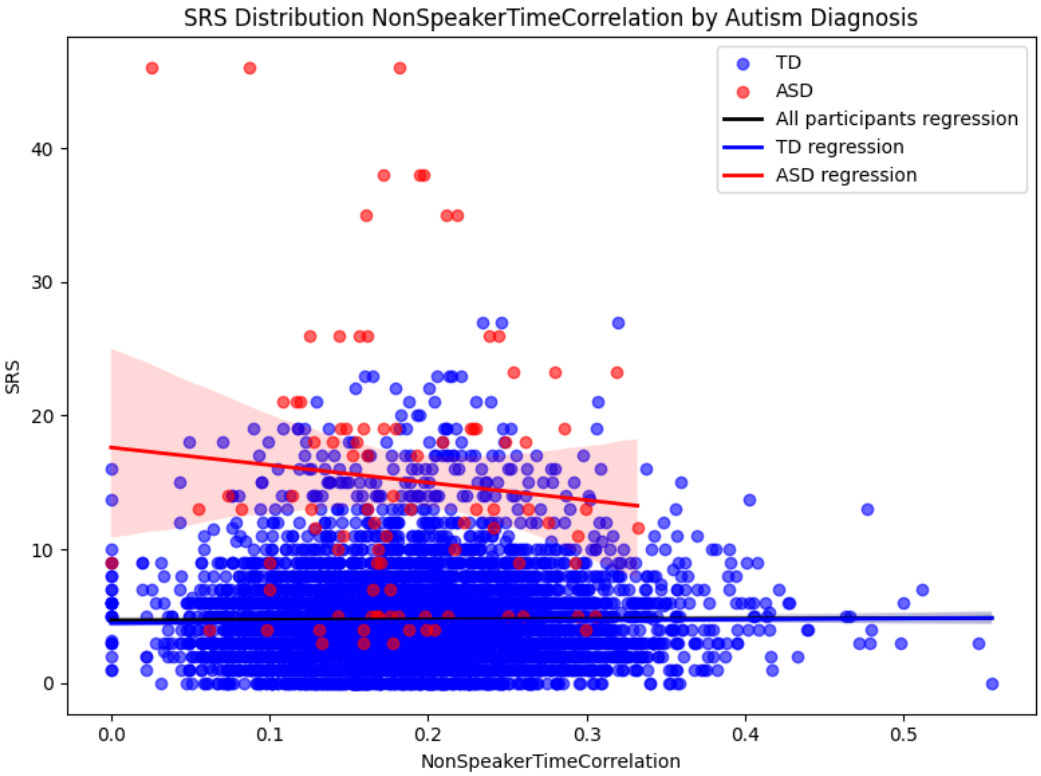


Figure B.28: Relationship between non-speaker time correlation and SRS scores by diagnosis.

Object Fixation

Table B.8: Log-likelihood and parameters for distribution fits on object fixation.

Participants	Distribution	LogLikelihood	μ	σ
All			0.02	0.04
	Normal	14191.60	0.02	0.00
	Beta	73102.06	0.01	0.00
	Gamma	**159089.48**	0.24	0.30
	Uniform	2774.63	0.35	0.04

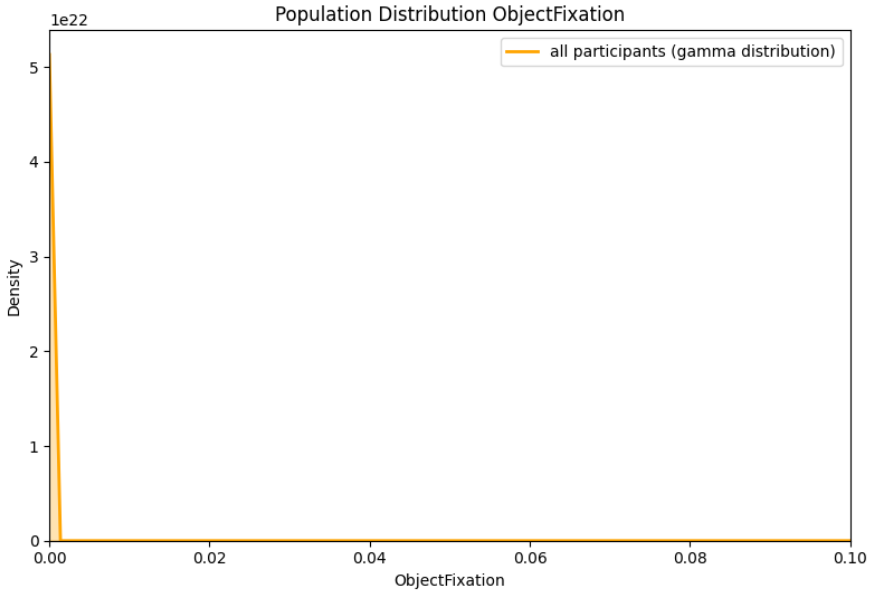


Figure B.29: Population distribution of object fixation.

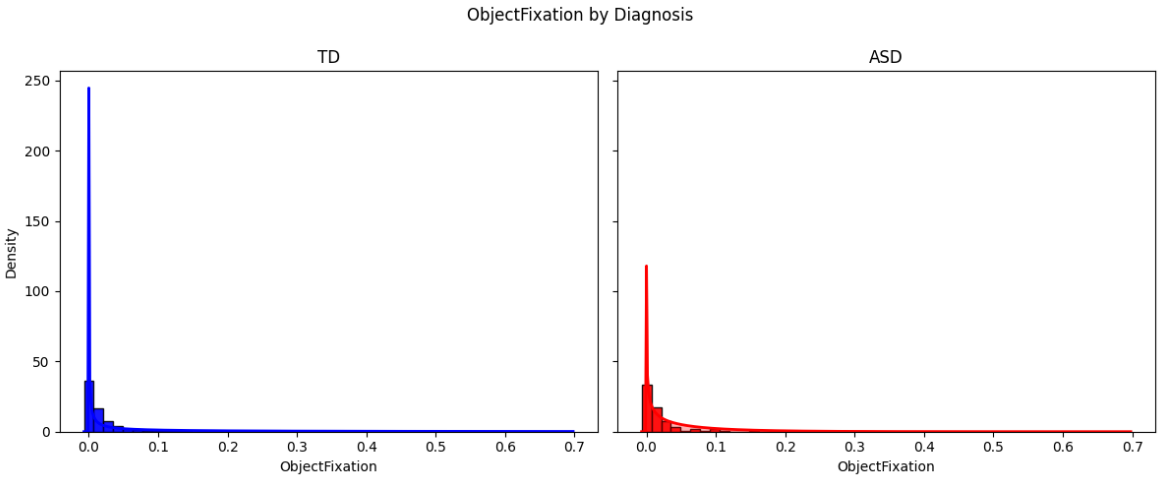


Figure B.30: Distribution by autism diagnosis, showing fitted distributions for each group.

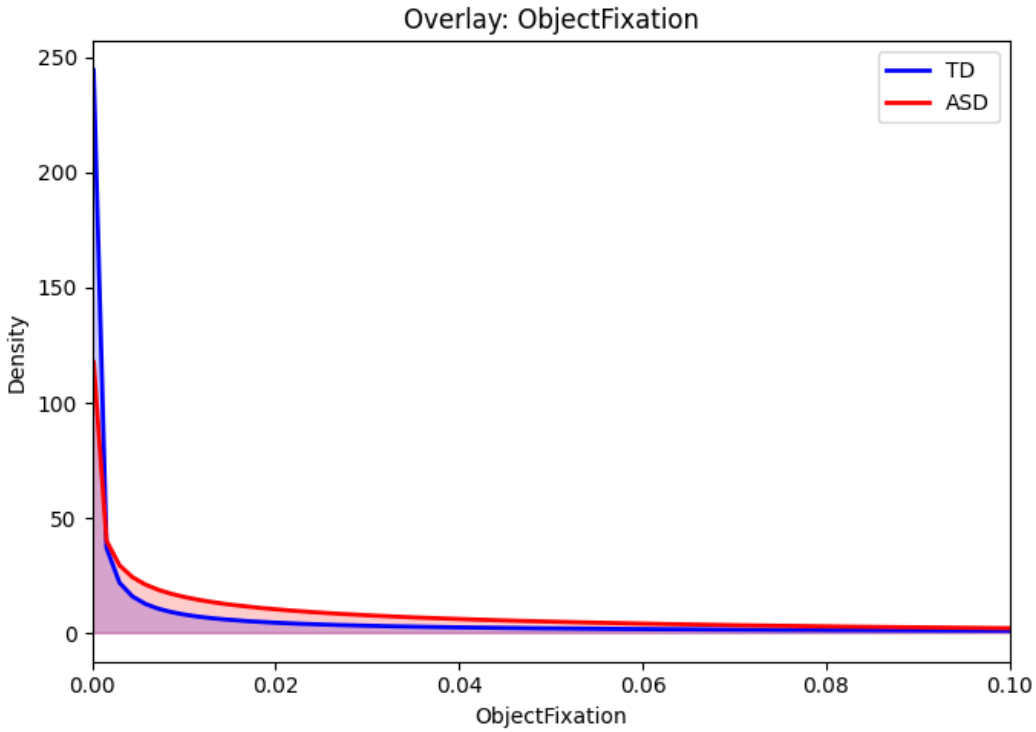


Figure B.31: Distribution overlays by autism diagnosis, showing fitted distributions for each group.

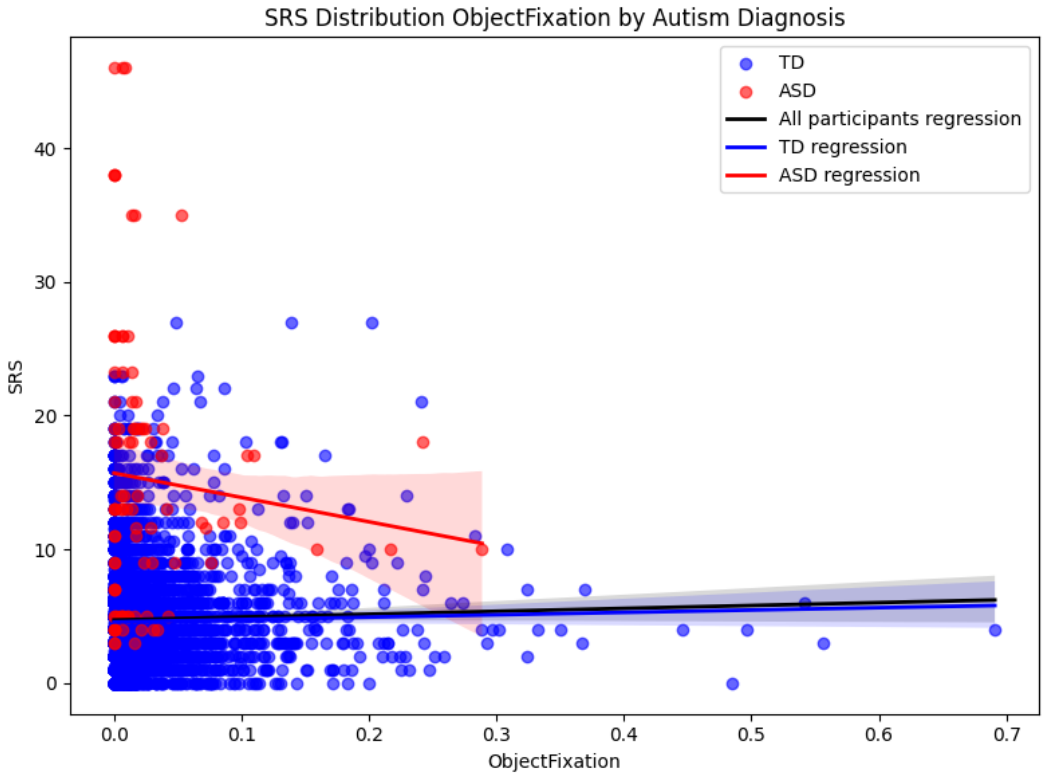


Figure B.32: Relationship between object fixation and SRS scores by diagnosis.

Predictive Saccades

Table B.9: Log-likelihood and parameters for distribution fits on predictive saccades.

Participants	Distribution	LogLikelihood	μ	σ
All			3.76	1.14
	Binominal	**-11108.38**	3.76	0.97
	Poisson	-13318.70	3.76	1.94

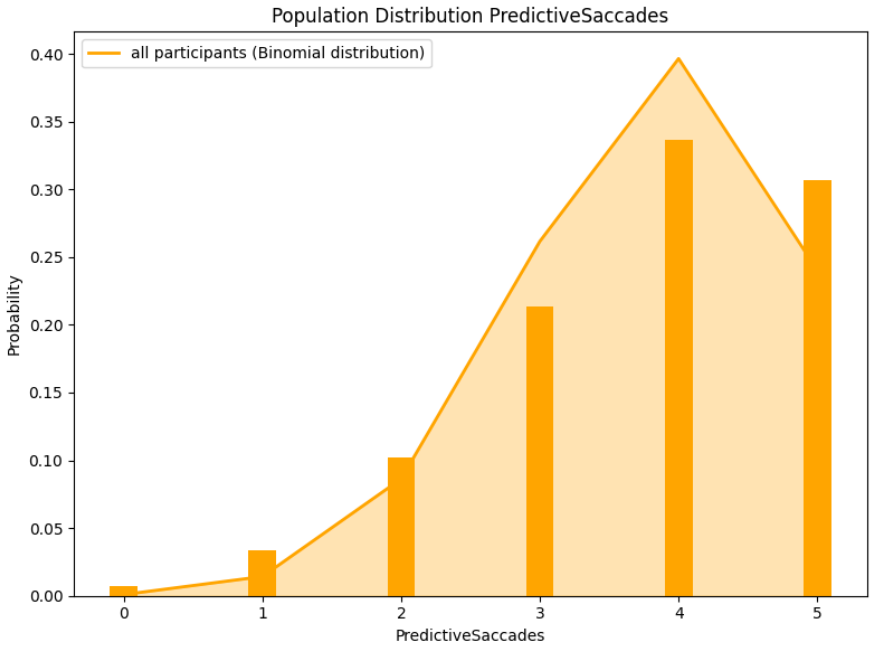


Figure B.33: Population distribution of predictive saccades.

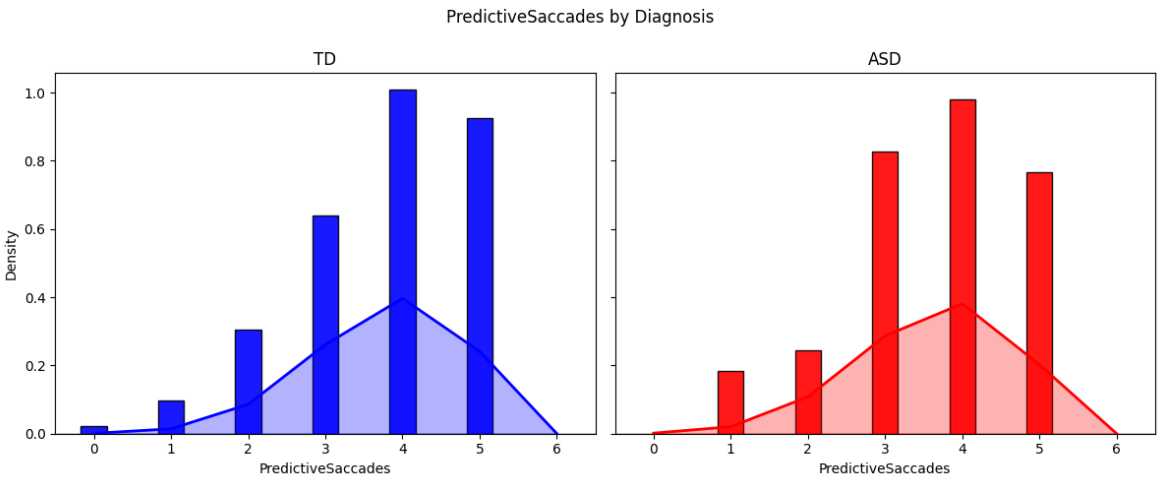


Figure B.34: Distribution by autism diagnosis, showing fitted distributions for each group.

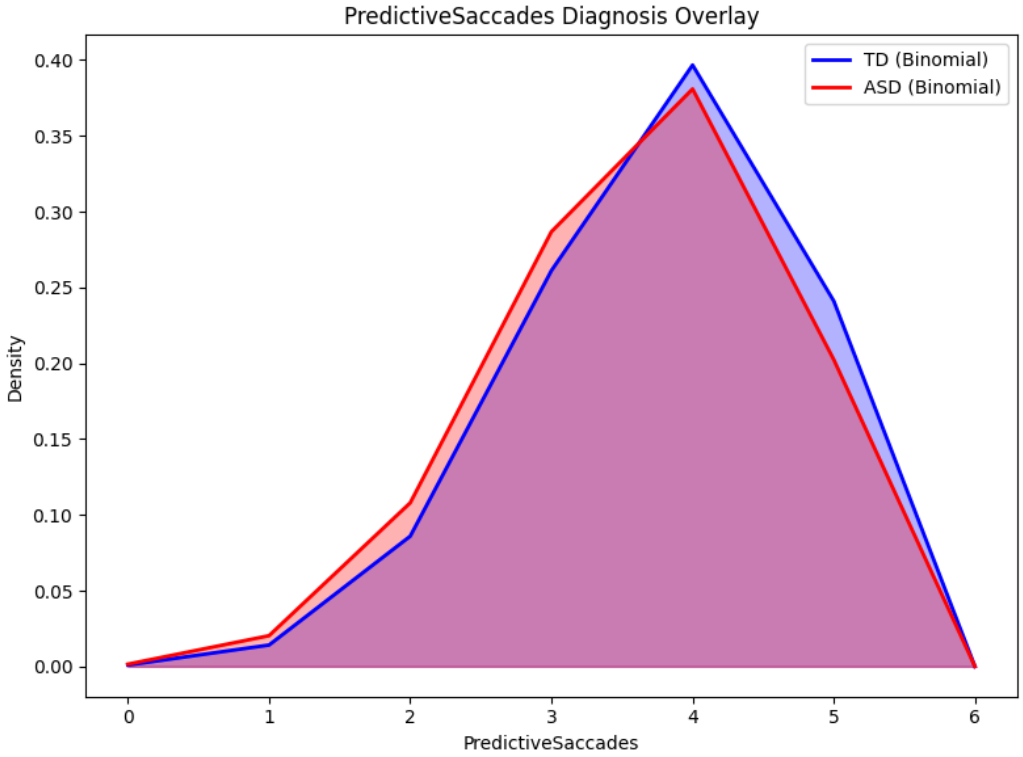


Figure B.35: Distribution overlays by autism diagnosis, showing fitted distributions for each group.

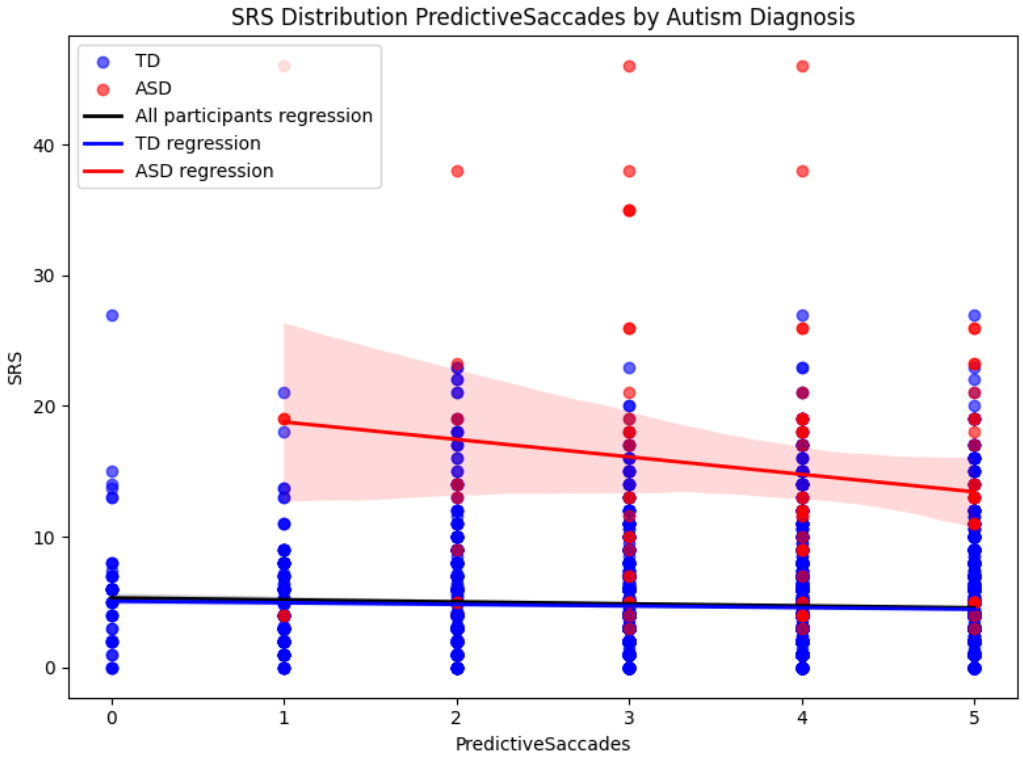


Figure B.36: Relationship between predictive saccades and SRS scores by diagnosis.

Saccadic Speed

Table B.10: Log-likelihood and parameters for distribution fits on saccadic speed.

Participants	Distribution	LogLikelihood	μ	σ
All			68.92	13.15
	Normal	**-29997.07**	68.92	172.92
	Gamma	-30100.03	68.71	180.23
	Uniform	-34745.81	65.86	871.98

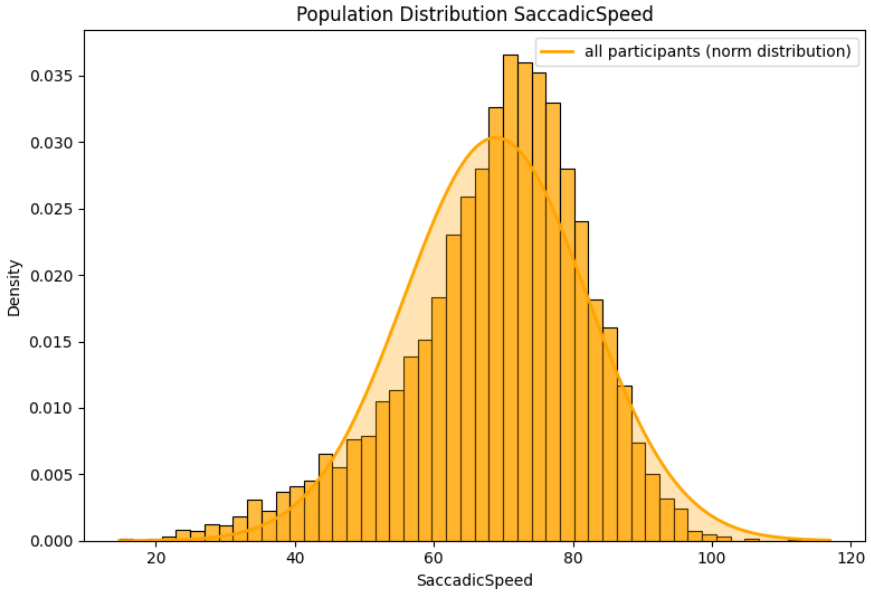


Figure B.37: Population distribution of saccadic speed.

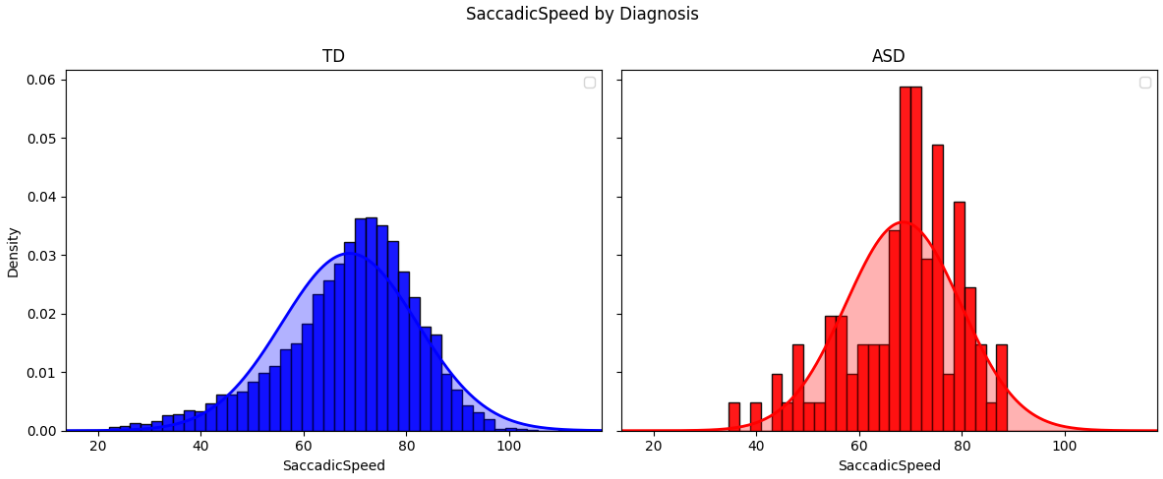


Figure B.38: Distribution by autism diagnosis, showing fitted distributions for each group.

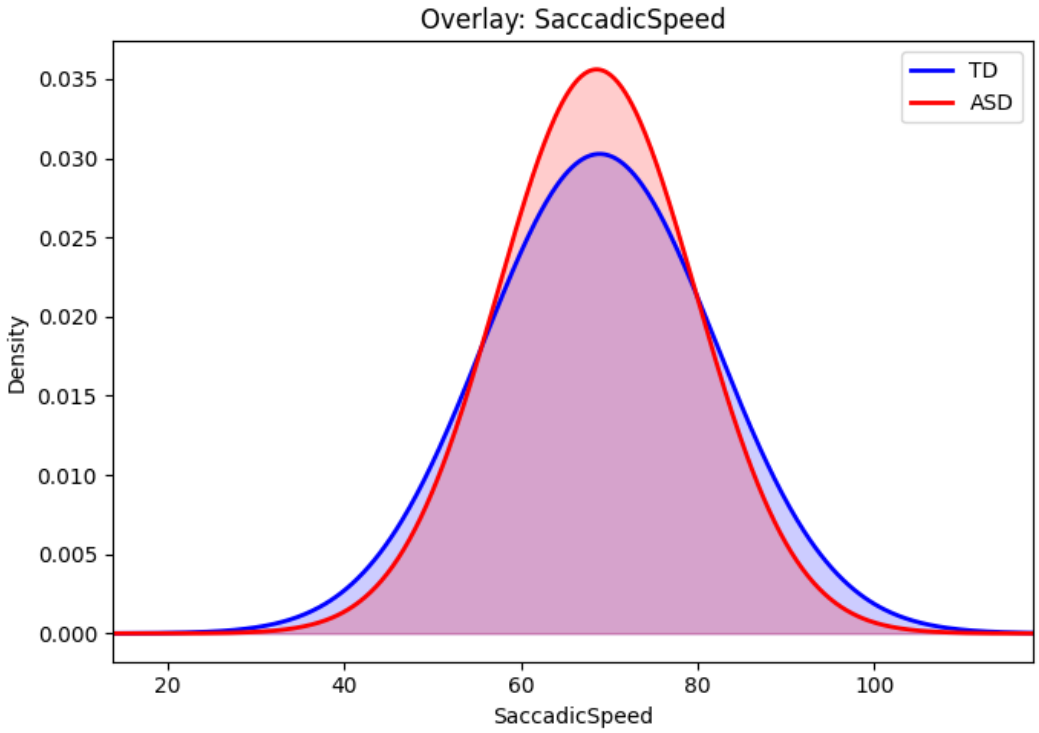


Figure B.39: Distribution overlays by autism diagnosis, showing fitted distributions for each group.

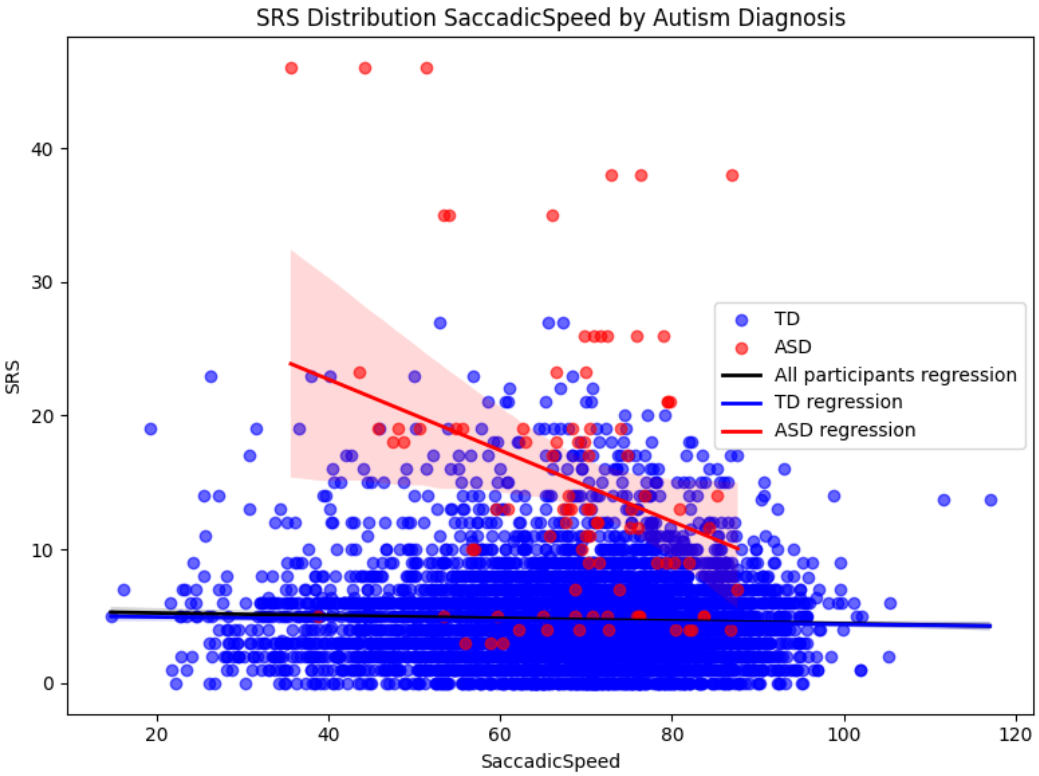


Figure B.40: Relationship between saccadic speed and SRS scores by diagnosis.

Scanpath Variance

Table B.11: Log-likelihood and parameters for distribution fits on scanpath variance.

Participants	Distribution	LogLikelihood	μ	σ
All			182.76	62.55
	Normal	-41706.43	182.76	3912.72
	Uniform	-50682.42	487.74	60840.17
	Gamma	**-40914.42**	182.76	3721.26

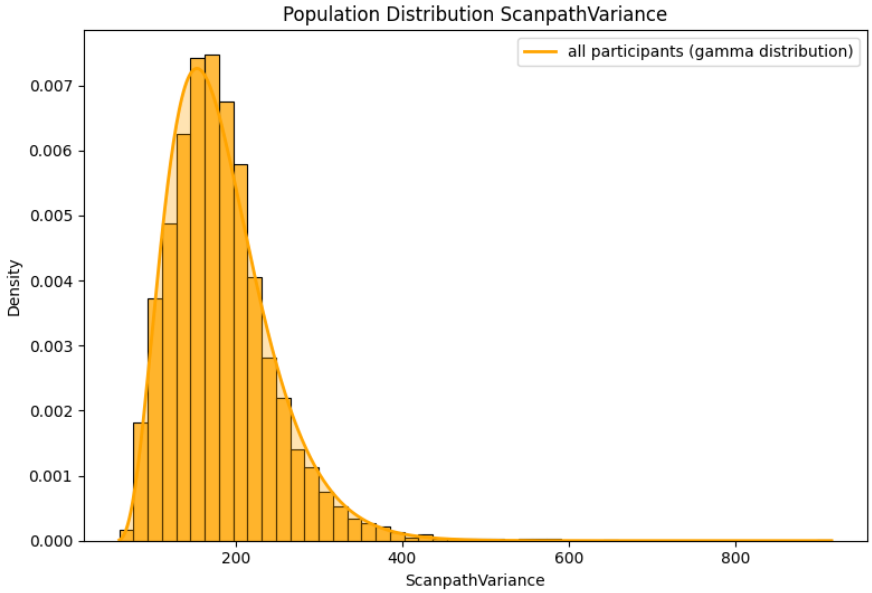


Figure B.41: Population distribution of scanpath variance.

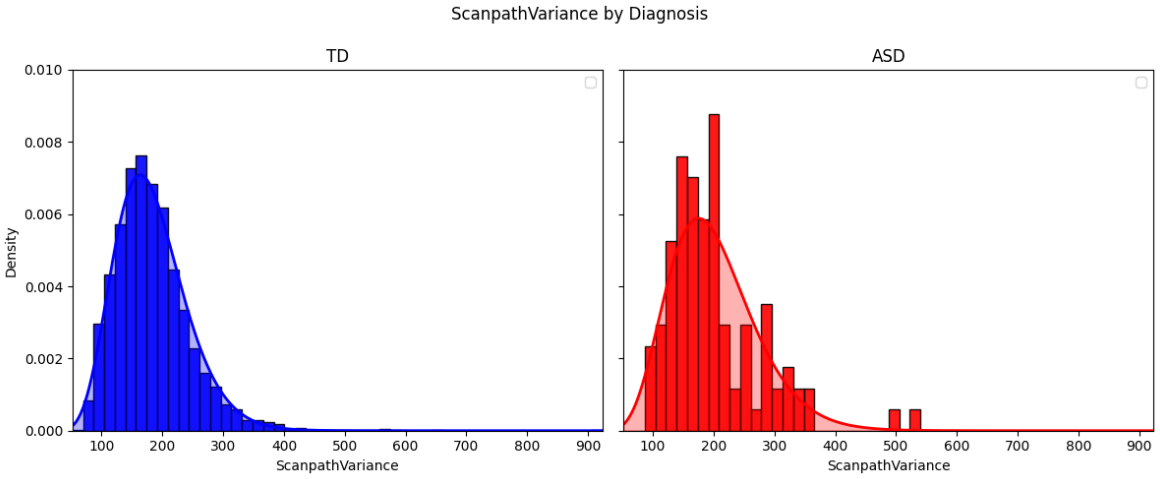


Figure B.42: Distribution by autism diagnosis, showing fitted distributions for each group.

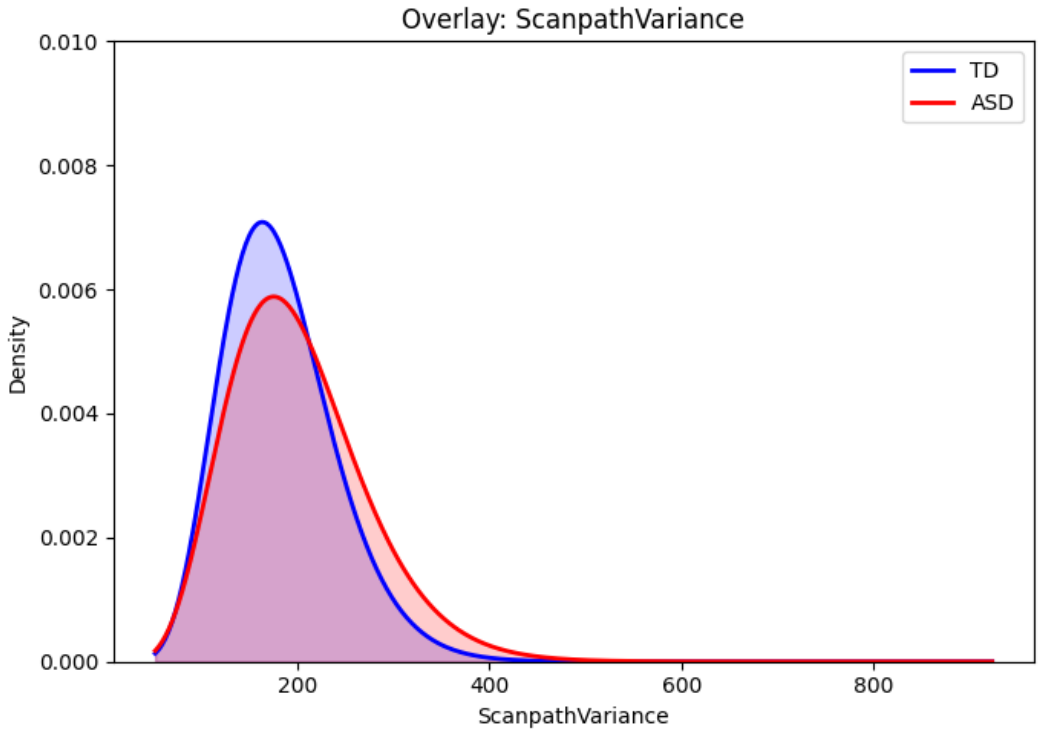


Figure B.43: Distribution overlays by autism diagnosis, showing fitted distributions for each group.

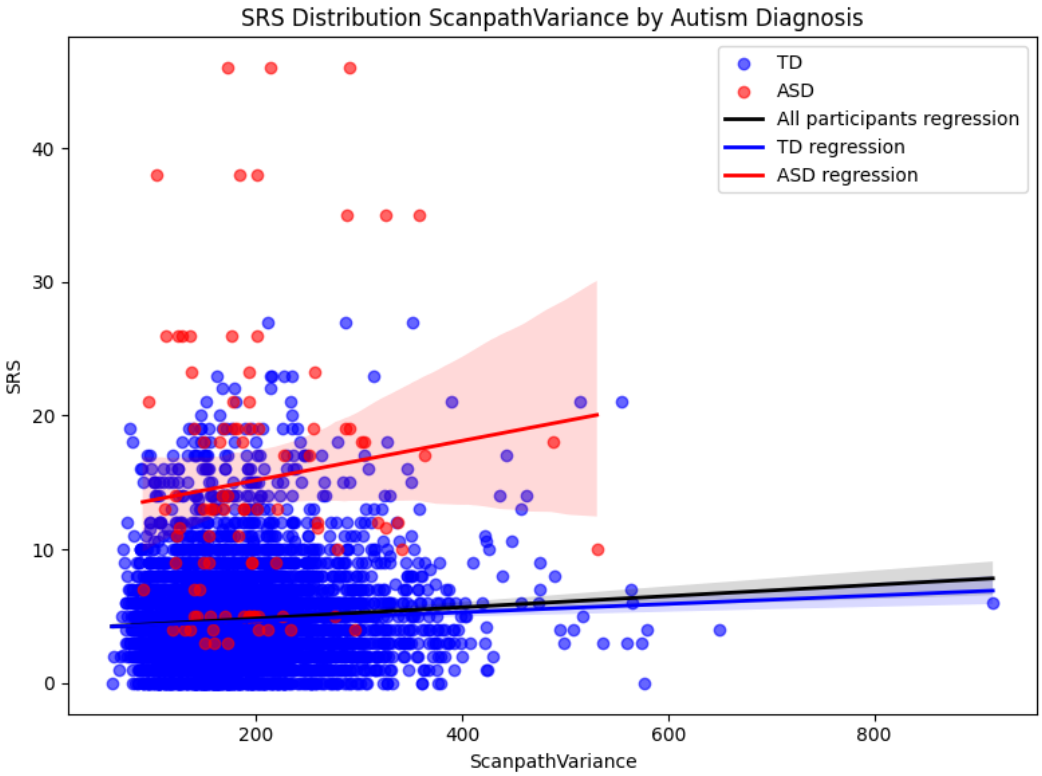


Figure B.44: Relationship between scanpath variance and SRS scores by diagnosis.

Screen Time

Table B.12: Log-likelihood and parameters for distribution fits on screen time.

Participants	Distribution	LogLikelihood	μ	σ
All			0.88	0.13
	Normal	4808.08	0.88	0.02
	Beta	**8851.98**	**0.88**	**0.01**
	Gamma	4449.76	0.88	0.02
	Uniform	2210.61	0.62	0.05

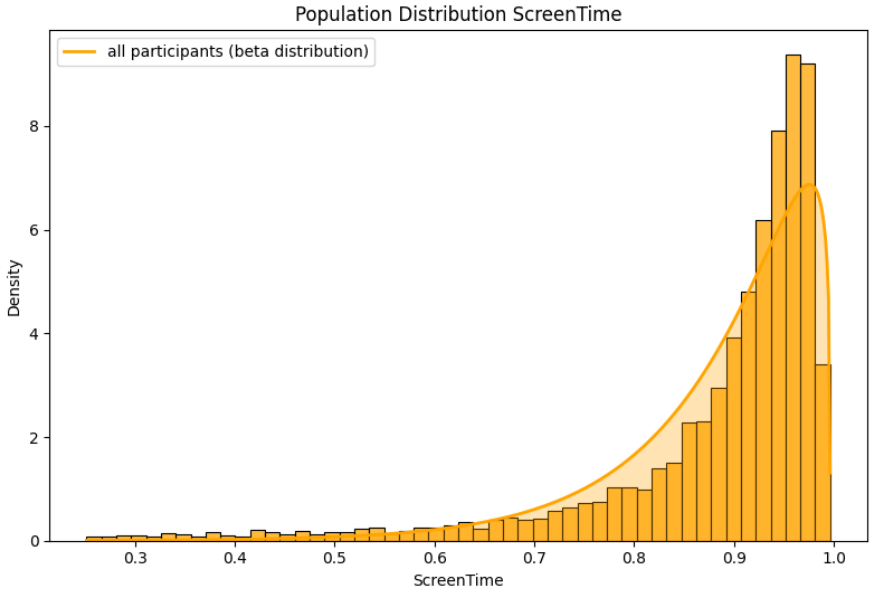


Figure B.45: Population distribution of screentime.

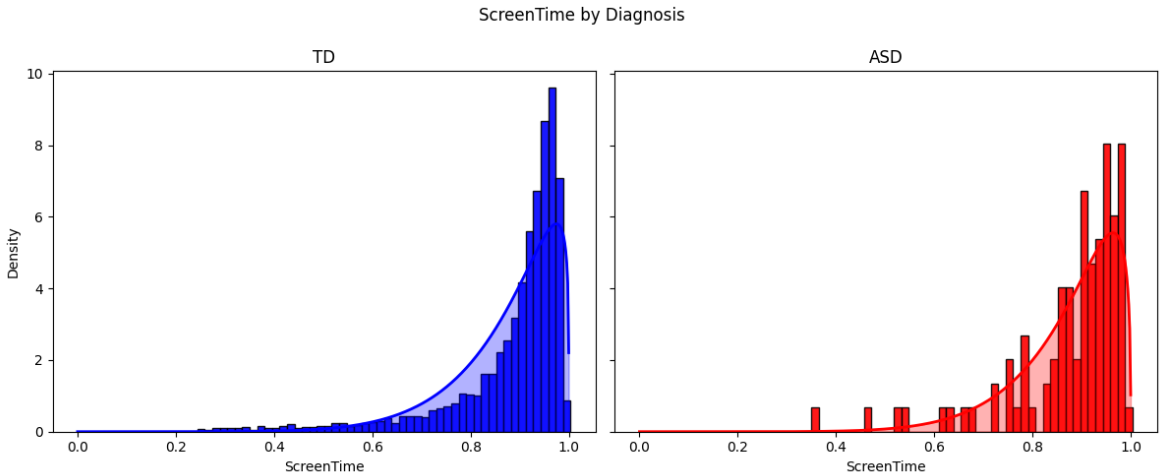


Figure B.46: Distribution by autism diagnosis, showing fitted distributions for each group.

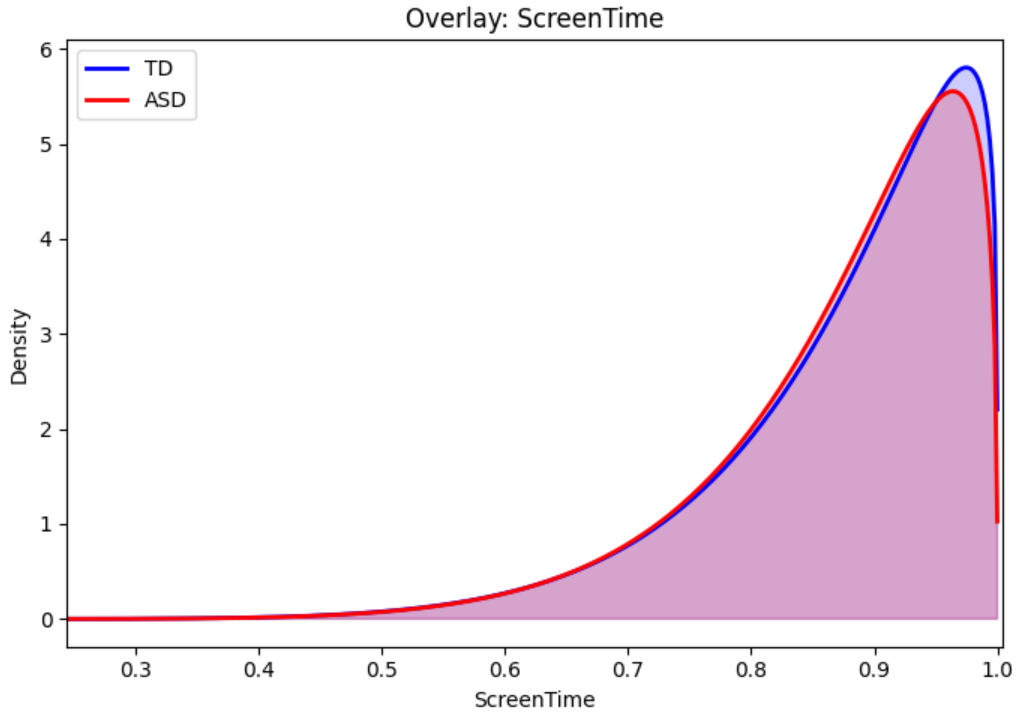


Figure B.47: Distribution overlays by autism diagnosis, showing fitted distributions for each group.

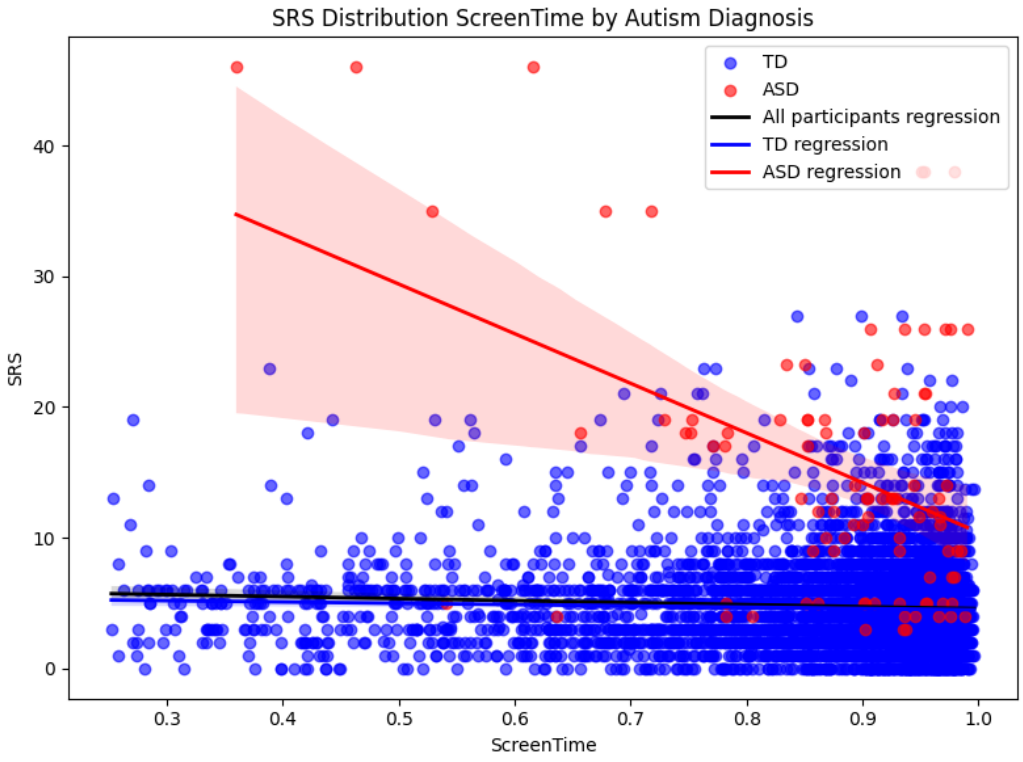


Figure B.48: Relationship between screentime and SRS scores by diagnosis.

Speaker Silhouette Correlation

Table B.13: Log-likelihood and parameters for distribution fits on speaker silhouette correlation.

Participants	Distribution	LogLikelihood	μ	σ
All			0.96	0.03
	Normal	15170.58	0.96	0.00
	Beta	**16842.45**	0.96	0.00
	Gamma	13928.42	0.96	0.00
	Uniform	3164.48	0.67	0.04

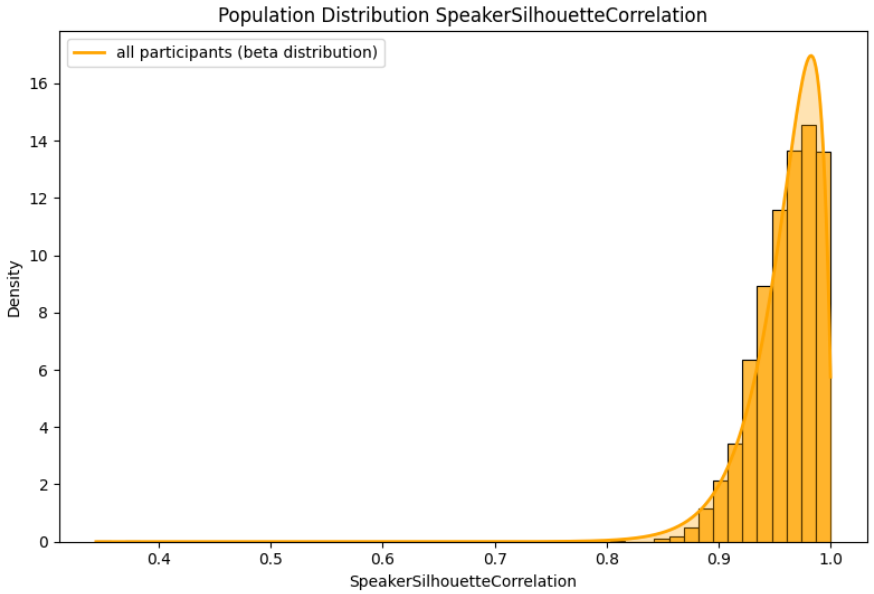


Figure B.49: Population distribution of speaker silhouette correlation.

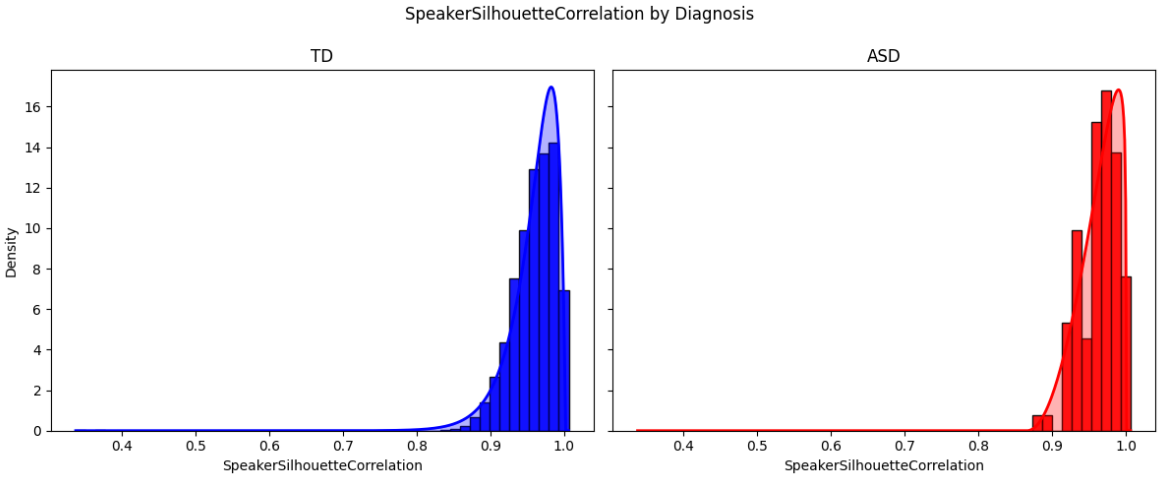


Figure B.50: Distribution by autism diagnosis, showing fitted distributions for each group.

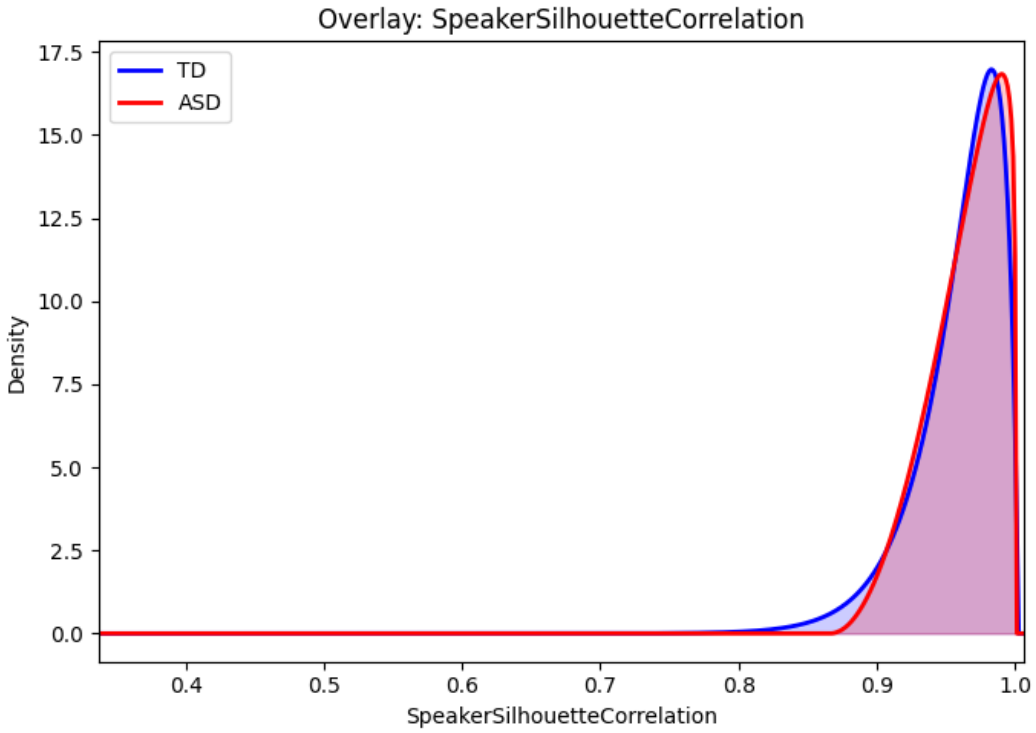


Figure B.51: Distribution overlays by autism diagnosis, showing fitted distributions for each group.

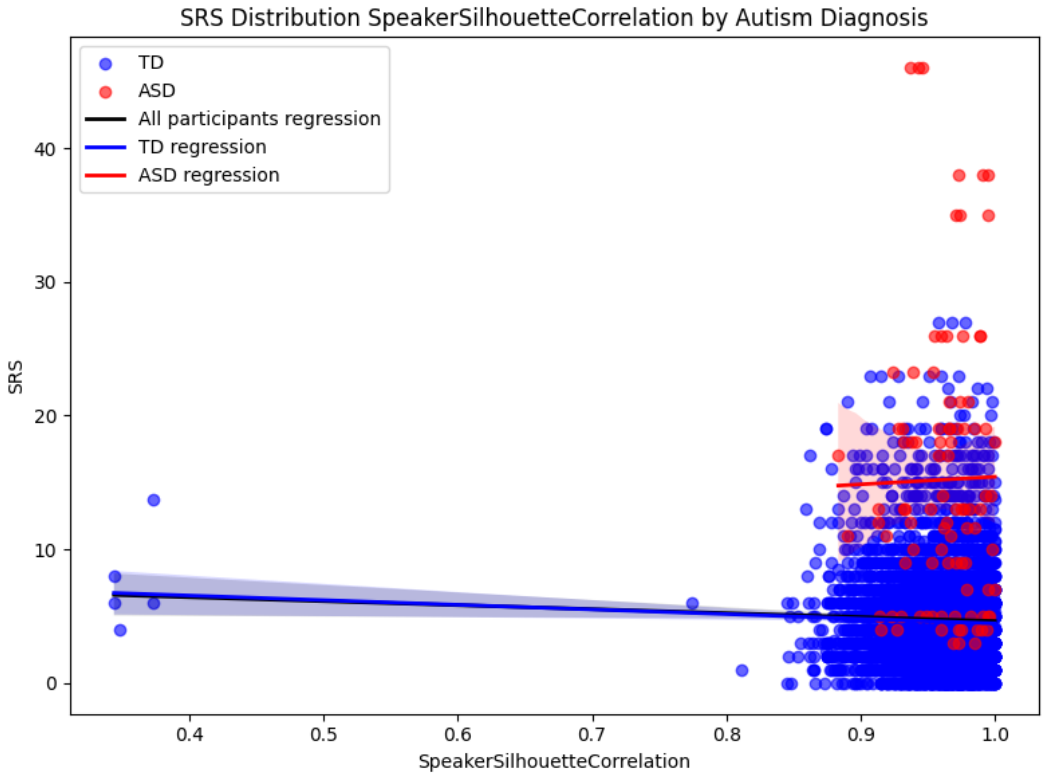


Figure B.52: Relationship between speaker silhouette correlation and SRS scores by diagnosis.

Speaker Time Correlation

Table B.14: Log-likelihood and parameters for distribution fits on speaker time correlation.

Participants	Distribution	LogLikelihood	μ	σ
All			0.60	0.12
	Normal	5432.63	0.60	0.01
	Beta	**5574.64**	0.60	0.12
	Gamma	5218.90	0.60	0.02
	Uniform	948.32	0.44	0.06

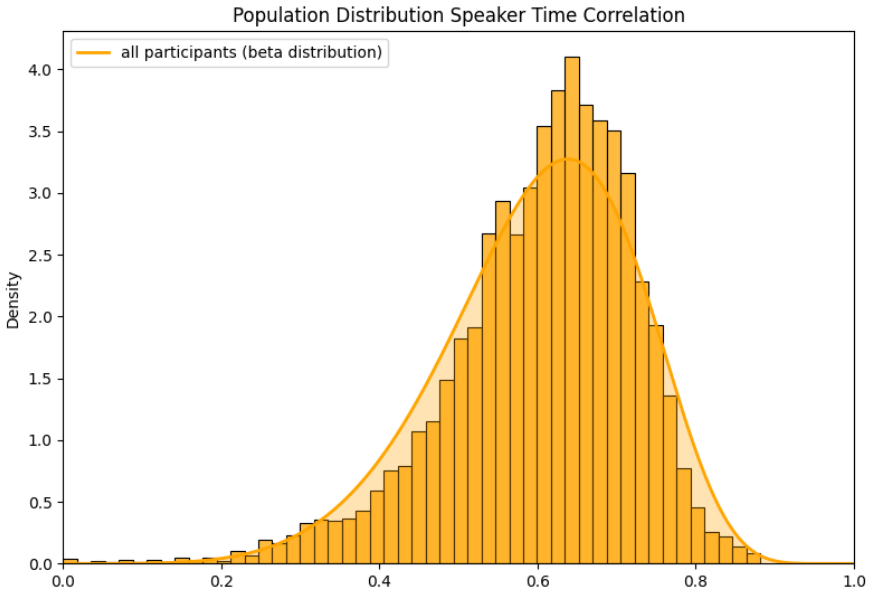


Figure B.53: Population distribution of speaker time correlation.

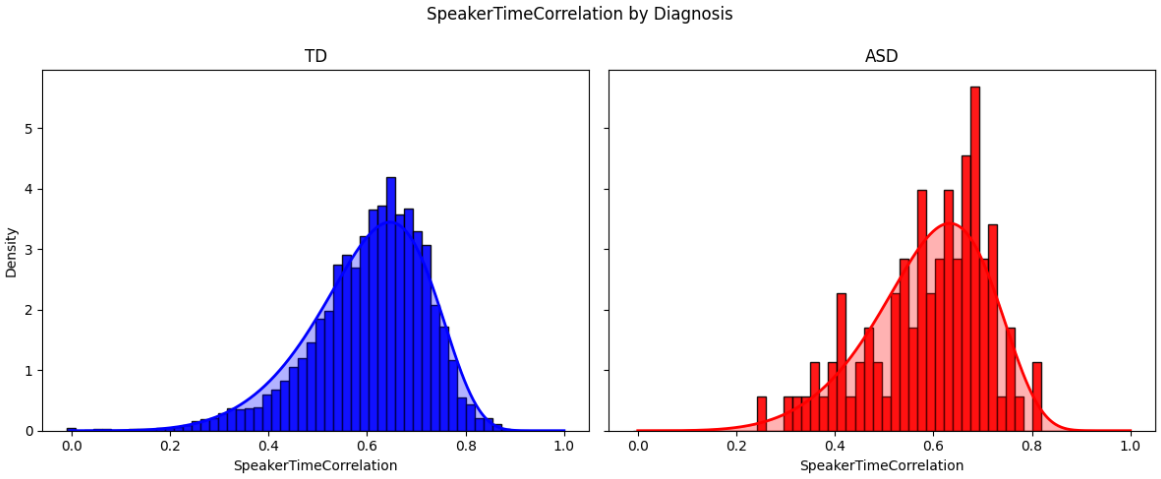


Figure B.54: Distribution by autism diagnosis, showing fitted distributions for each group.

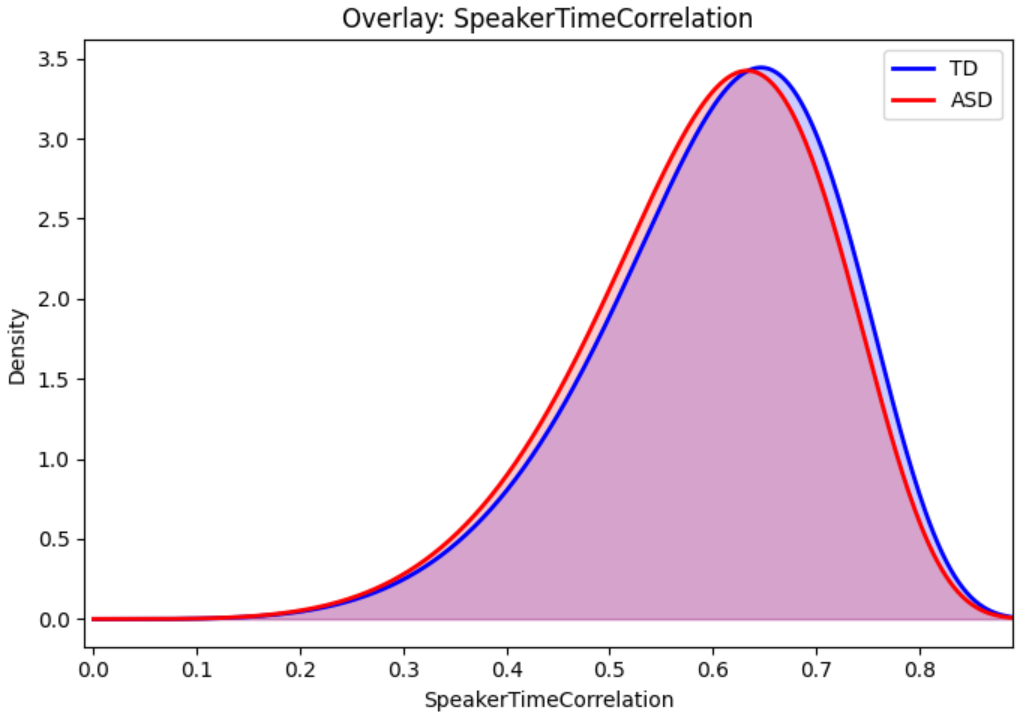


Figure B.55: Distribution overlays by autism diagnosis, showing fitted distributions for each group.

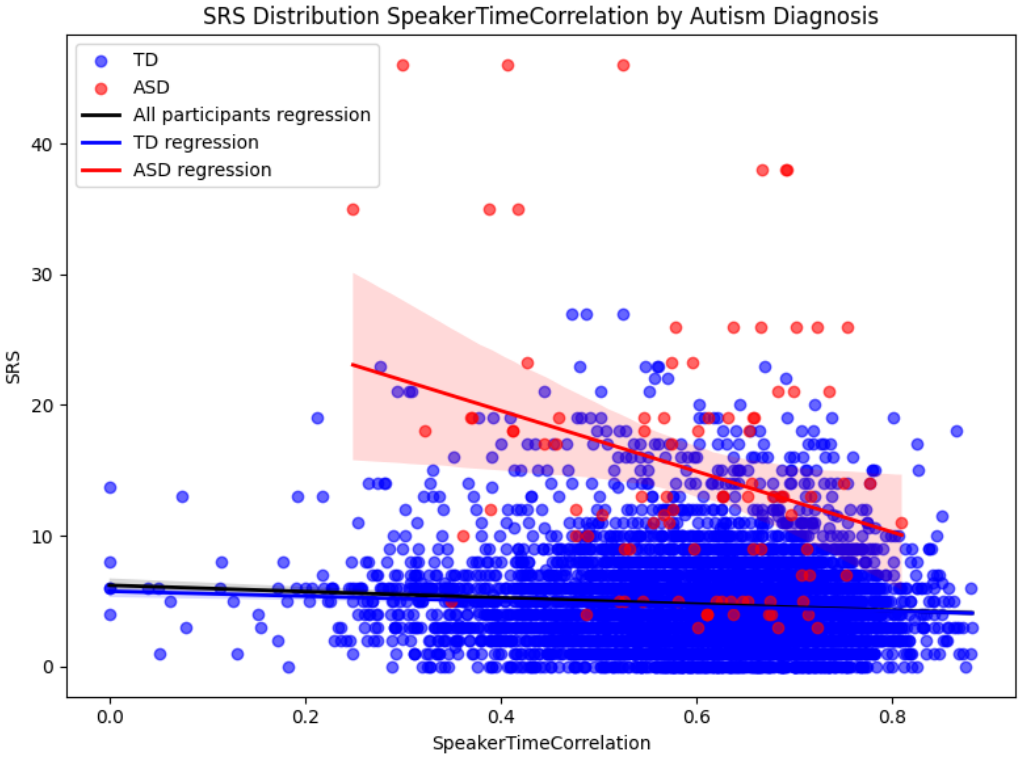


Figure B.56: Relationship between speaker time correlation and SRS scores by diagnosis.

Time Per Fixation

Table B.15: Log-likelihood and parameters for distribution fits on time per fixation.

Participants	Distribution	LogLikelihood	μ	σ
All			475157.04	183371.09
	Normal	-101644.85	475157.04	33624957848.41
	Gamma	**-101632.31**	450658.78	32825528270.28
	Uniform	-108208.90	991824.78	275004203793.78

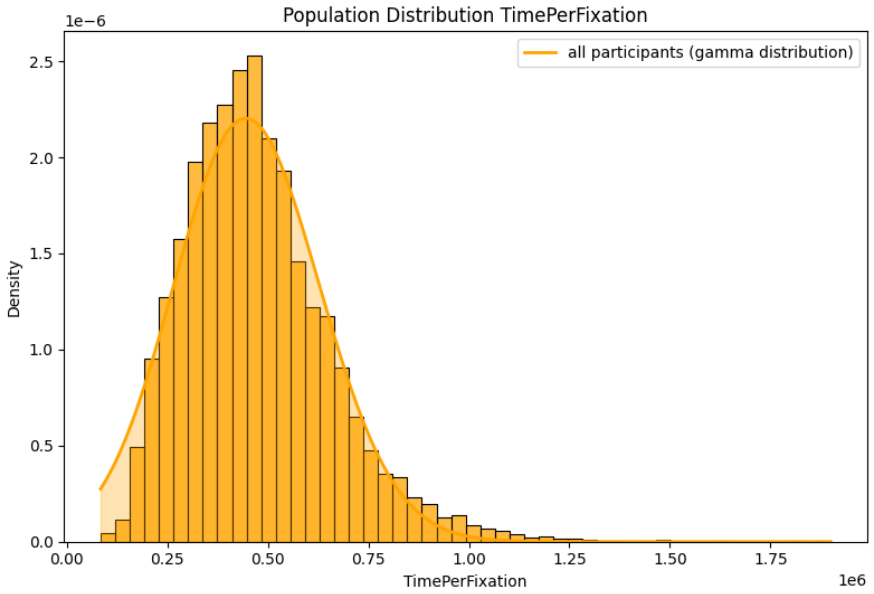


Figure B.57: Population distribution of time per fixation.

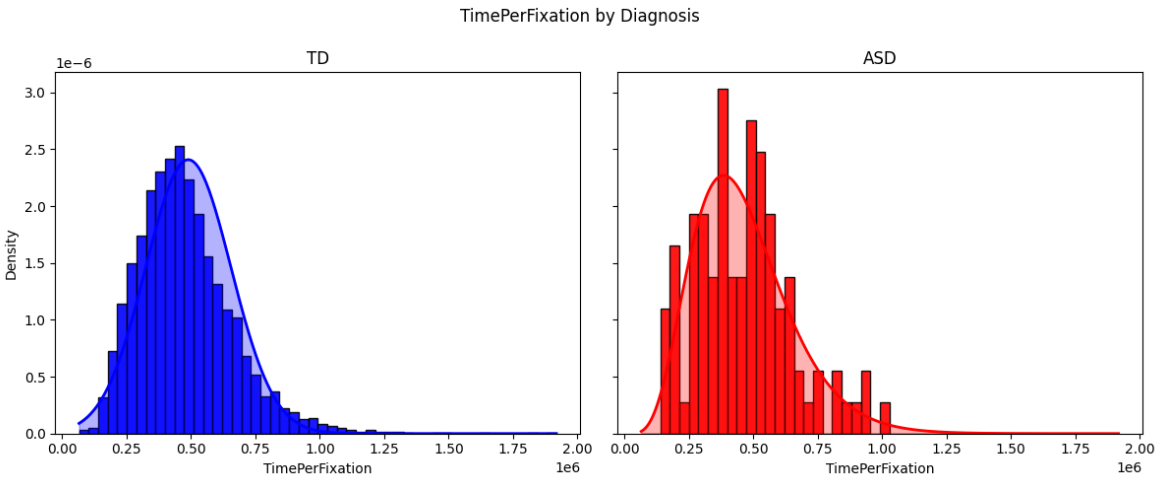


Figure B.58: Distribution by autism diagnosis, showing fitted distributions for each group.

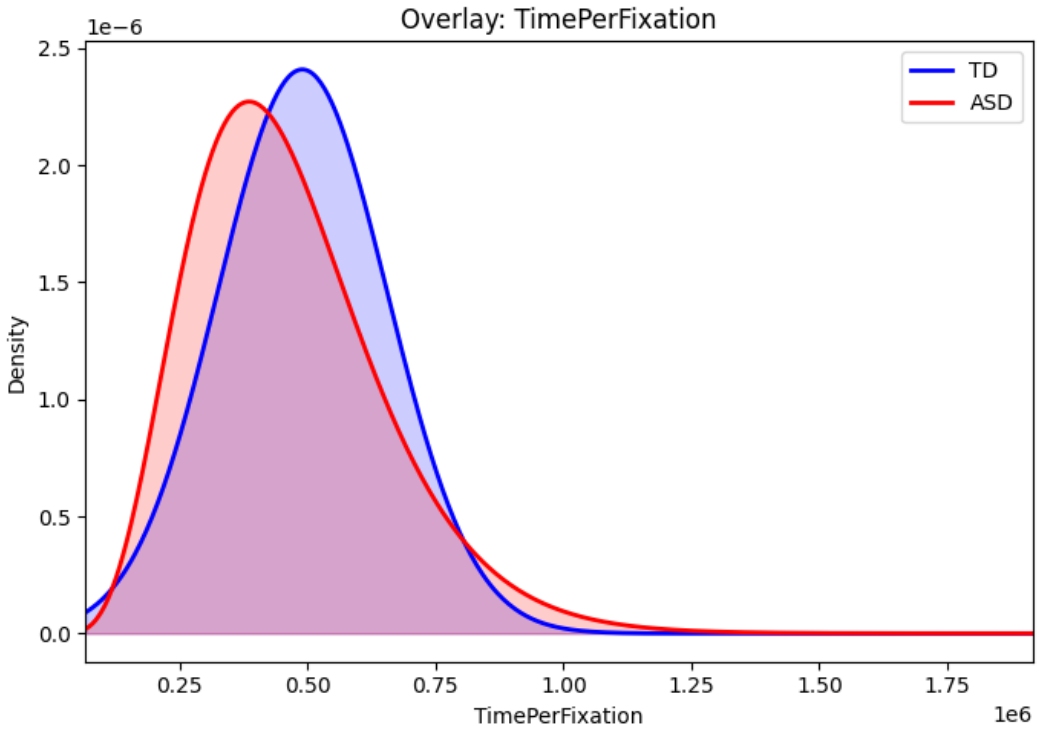


Figure B.59: Distribution overlays by autism diagnosis, showing fitted distributions for each group.

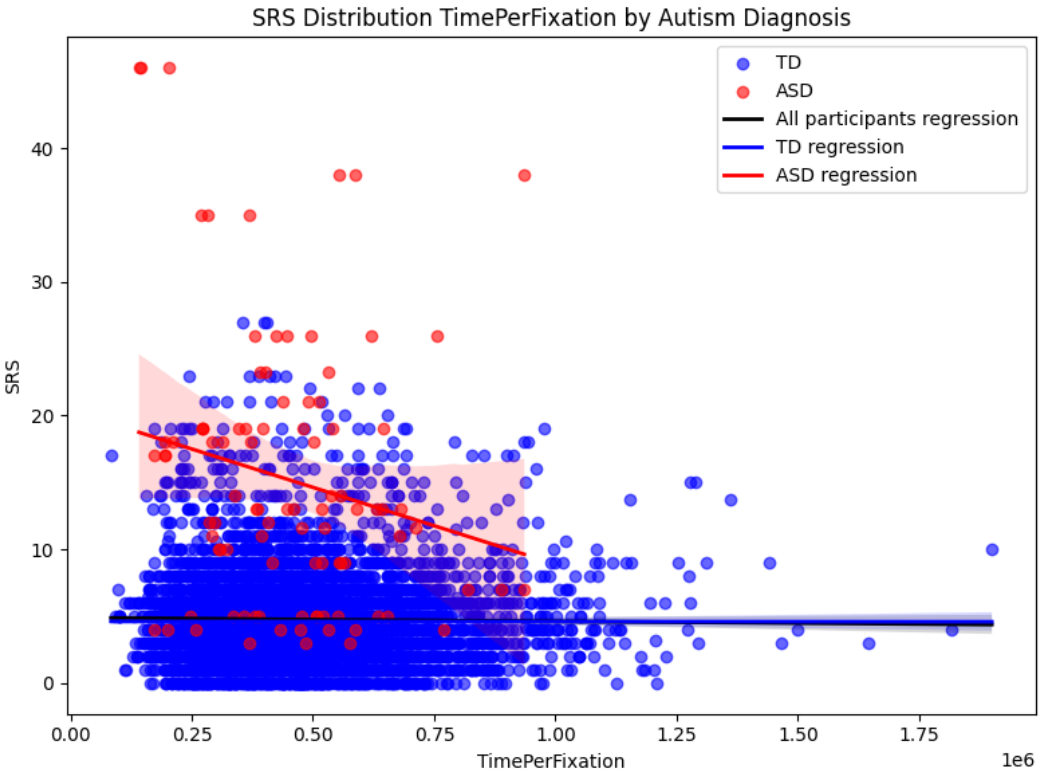


Figure B.60: Relationship between time per fixation and SRS scores by diagnosis.

C

Video-Related Effects

Background Fixation

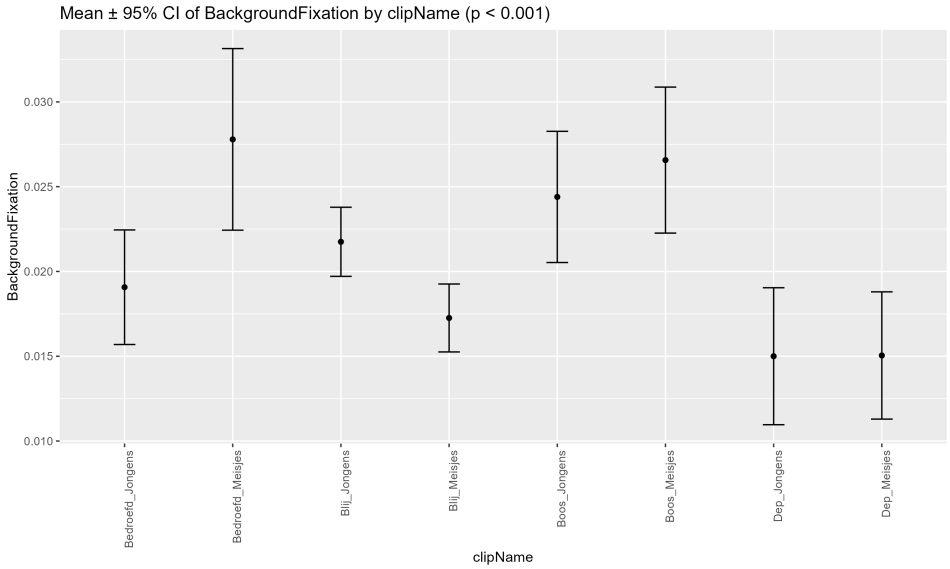


Figure C.1: Mean values with 95% CI per video clip for the fixations on the background. Variation across clips may reflect differences in how visually or socially engaging the conversations were. In less engaging clips, participants may have been more prone to disengage from the social scene and let their gaze drift toward the background.

Body Fixation

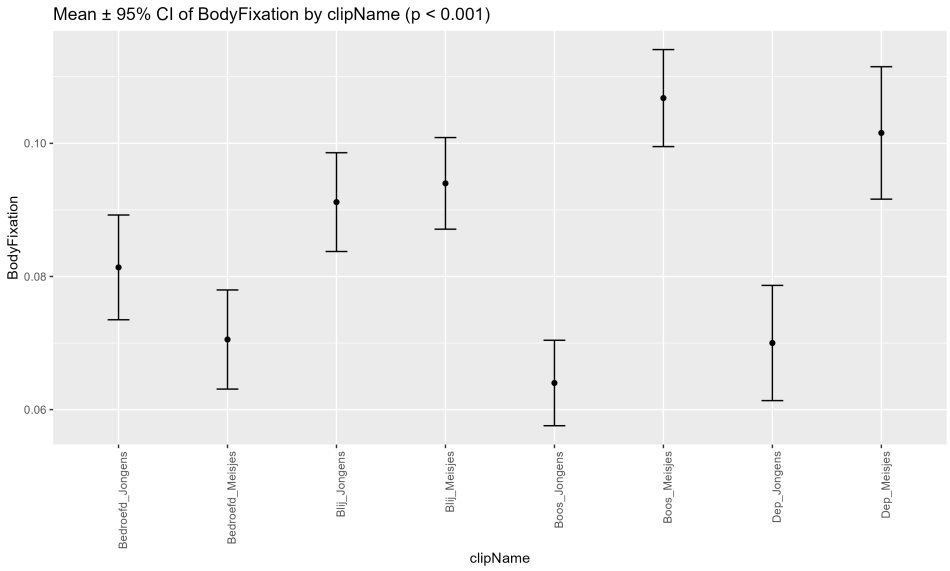


Figure C.2: Mean values with 95% CI per video clip for the fixations on the body, excluding the face, of characters. Differences across clips may be driven by variation in gesture frequency, bodily expressiveness, or visual framing.

Eyes Fixation

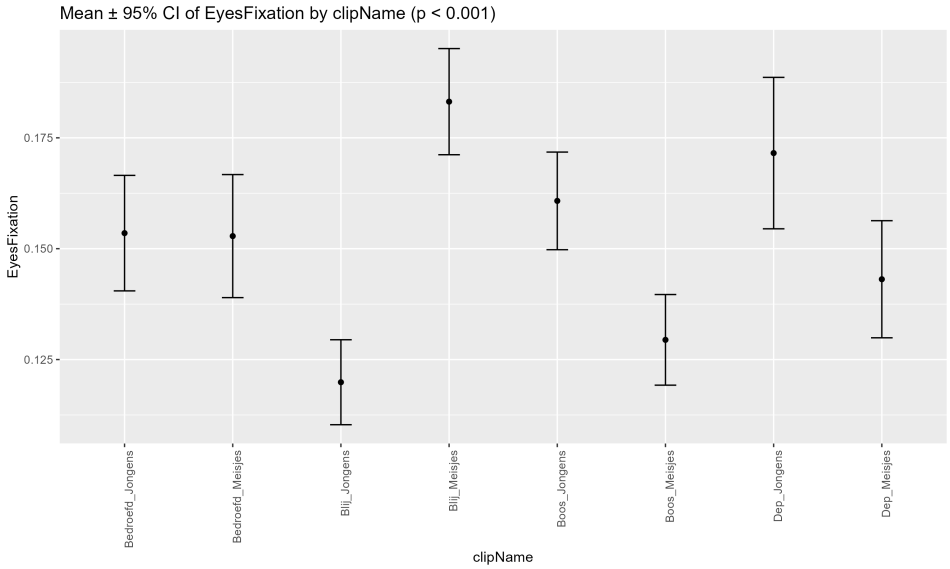


Figure C.3: Mean values with 95% CI per video clip for the fixations on the eyes. Variations across clips likely reflect differences in emotional expressiveness, eye visibility, or gaze dynamics of the characters.

Face Fixation

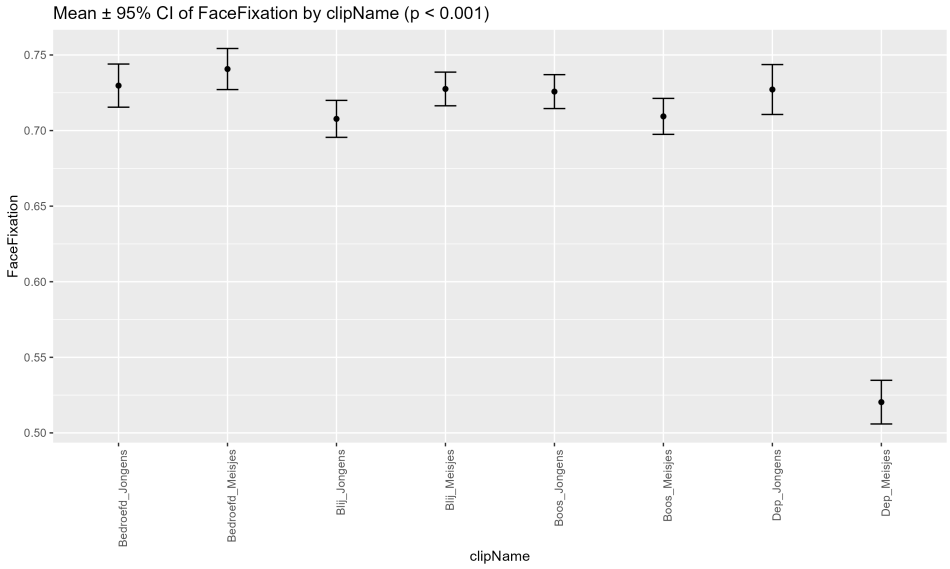


Figure C.4: Mean values with 95% CI per video clip for the fixations on the face. Variation across clips may result from differences in framing, or visual competition from other elements.

Mouth Fixation

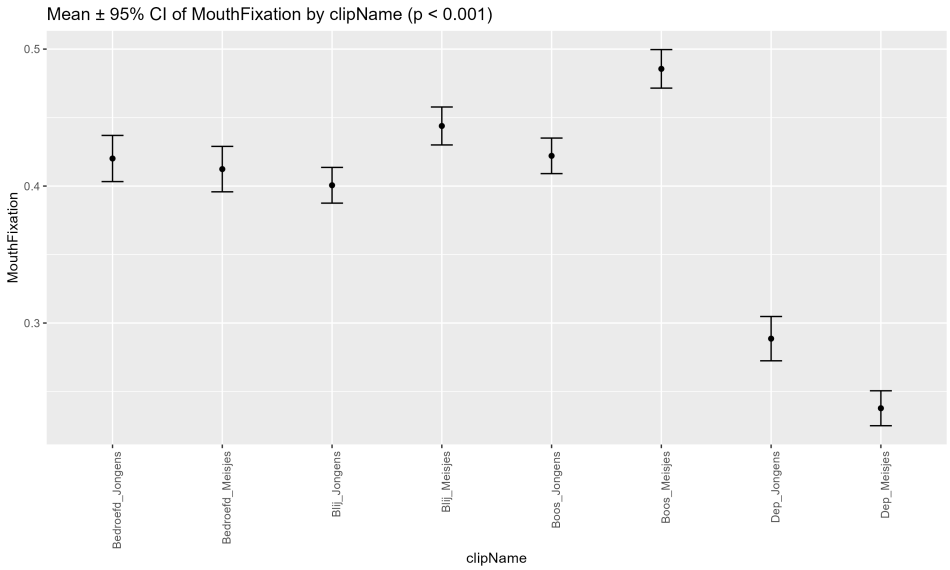


Figure C.5: Mean values with 95% CI per video clip for the fixations on the mouth. Variation between clips may reflect differences in speech articulation, visibility of mouth movements, or audiovisual clarity.

Non-Speaker Silhouette Correlation

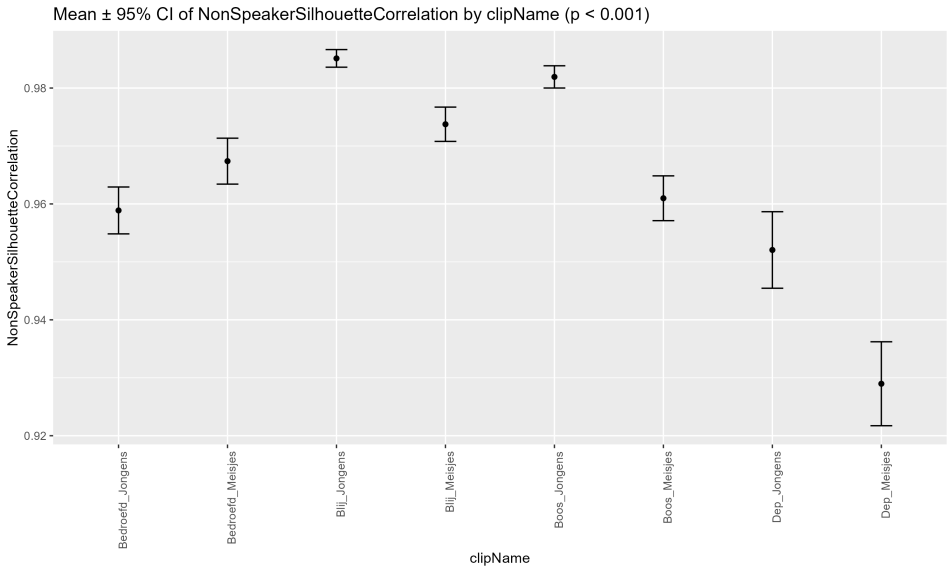


Figure C.6: Mean and 95% CI of Non-Speaker Silhouette Correlation across video clips. This metric assesses how structurally aligned participants' gaze patterns were with the non-speaking individuals during speech, focusing on the overall shape of the gaze sequence rather than precise timing. Variability across clips indicates that some non-speakers may have elicited more sustained or patterned attention, despite not speaking, possibly due to contextual or social cues.

Non-Speaker Time Correlation

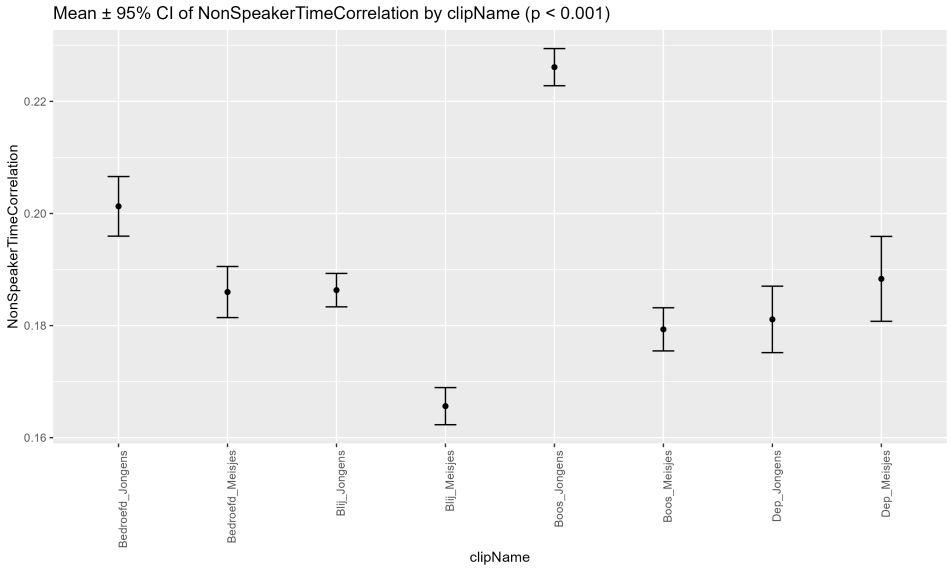


Figure C.7: Mean and 95% CI of Non-Speaker Time Correlation across video clips. This figure illustrates the degree to which gaze behaviour tracked non-speaking individuals during speech. Differences between clips suggest that certain non-speakers may have drawn more attention, potentially due to social dynamics, actions, or visual salience within the scene.

Object Fixation

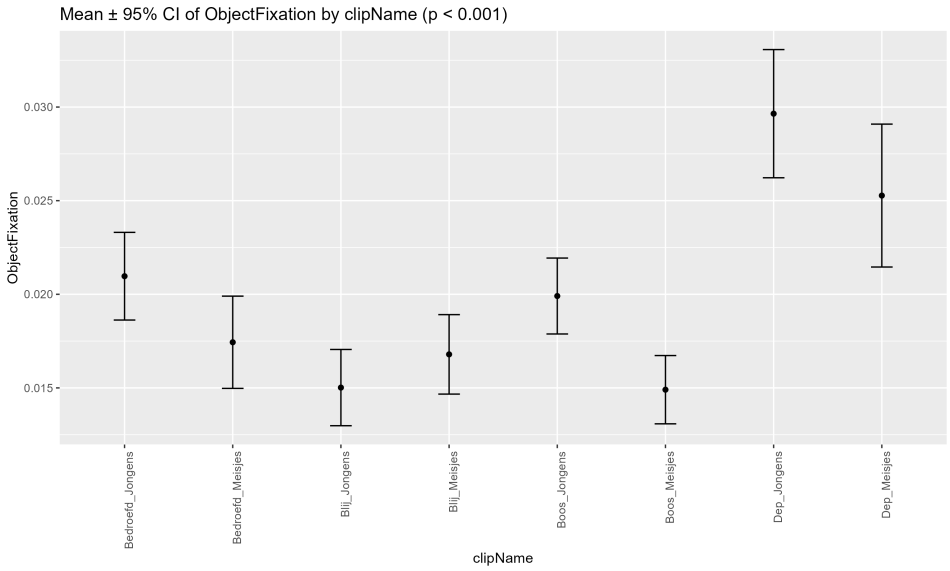


Figure C.8: Mean values with 95% CI per video clip for the fixations on the distracting object in the scene. Higher fixation rates on objects may indicate reduced social engagement. Variability across clips may be influenced by how captivating the conversation was.

Predictive Saccades

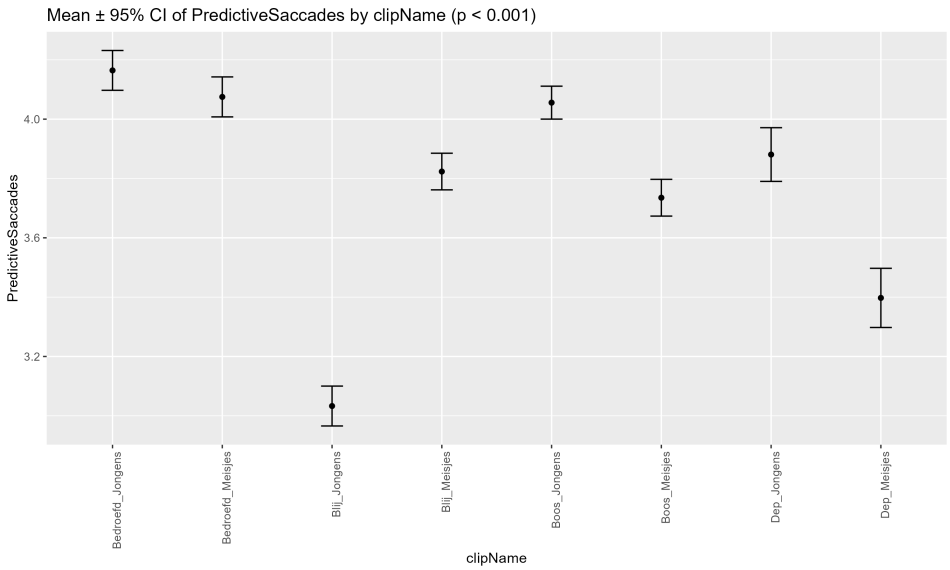


Figure C.9: Mean and 95% CI of Predictive Saccades across video clips. Predictive saccades refer to anticipatory eye movements made ahead of speaker turns. Differences in predictive behaviour across clips may indicate that some scenes offered more predictable visual or narrative structure, allowing participants to anticipate upcoming speech shifts more easily.

Saccadic Speed

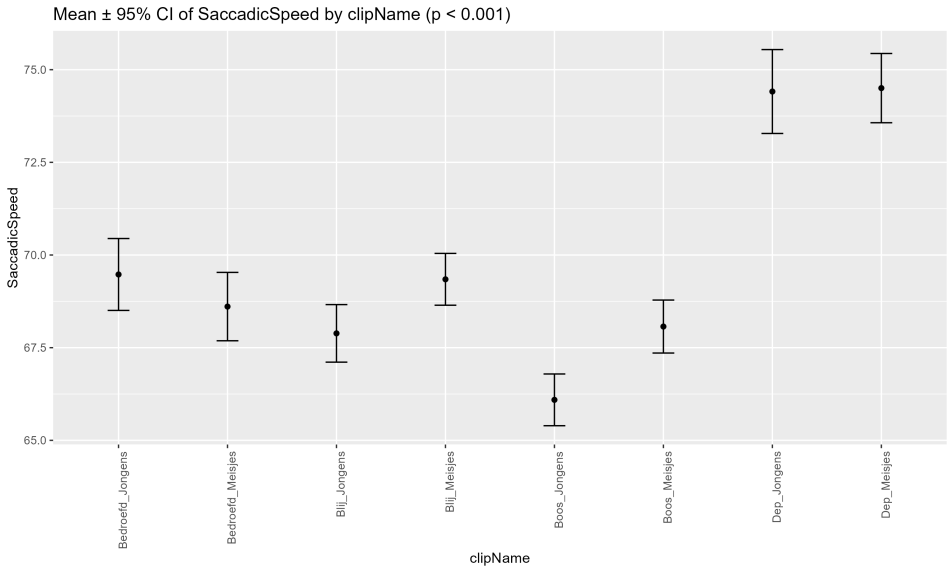


Figure C.10: Mean and 95% CI of Saccadic Speed across video clips. Saccadic speed reflects the average velocity of rapid eye movements between fixations. Variations across clips suggest that content influenced how quickly participants shifted their gaze, potentially due to differences in pacing, motion, or attentional demands.

Scanpath Variance

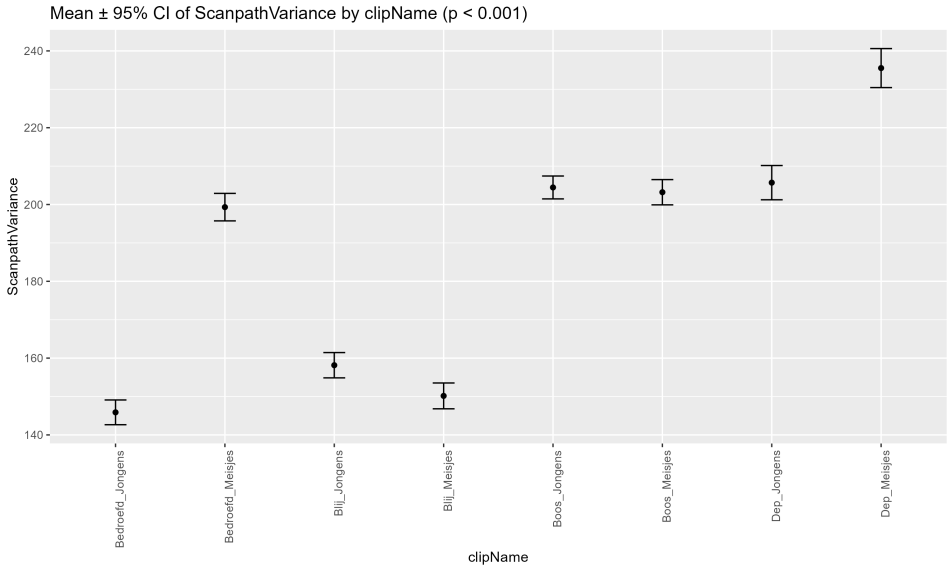


Figure C.11: Mean and 95% CI of Scanpath Variance across video clips. This metric quantifies how much the gaze position of an individual deviates from the average across all participants. The remaining video-related effects on this metric are most likely due to differences in the spatical layout across videos.

Screen Time

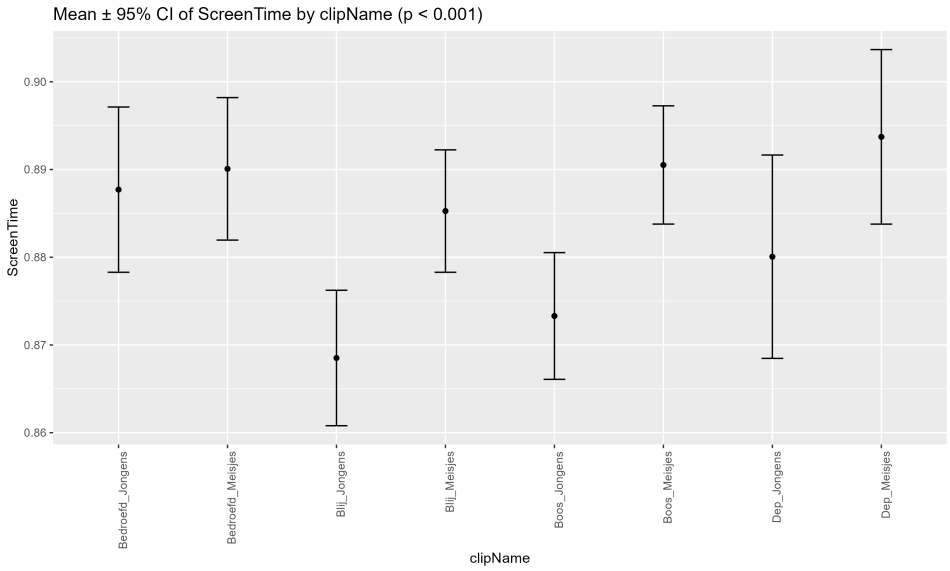


Figure C.12: Mean and 95% CI of Screen Time across video clips. This plot displays the average proportion of time participants spent looking at the screen during each video clip, with error bars indicating the 95% CI. Variations in Screen Time across clips reflect differences in how engaging or visually accessible the clips were, highlighting the influence of stimulus content—such as emotional salience or actor behaviour—on participants' viewing behaviour.

Speaker Silhouette Correlation

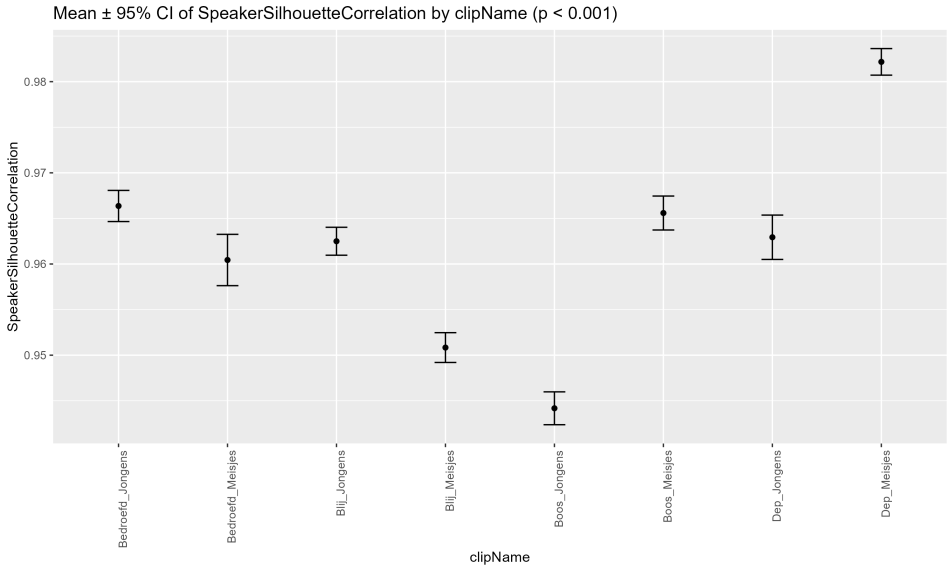


Figure C.13: Mean and 95% CI of Speaker Silhouette Correlation across video clips. This measure captures the structural similarity between the participant's gaze pattern and the speaker's speech activity, reflecting the shape and continuity of engagement rather than just temporal overlap. Unlike time correlation, silhouette correlation penalises fragmented viewing behaviour, such as brief glances. Differences across clips may reflect variations in how continuously participants tracked the speaker during speech.

Speaker Time Correlation

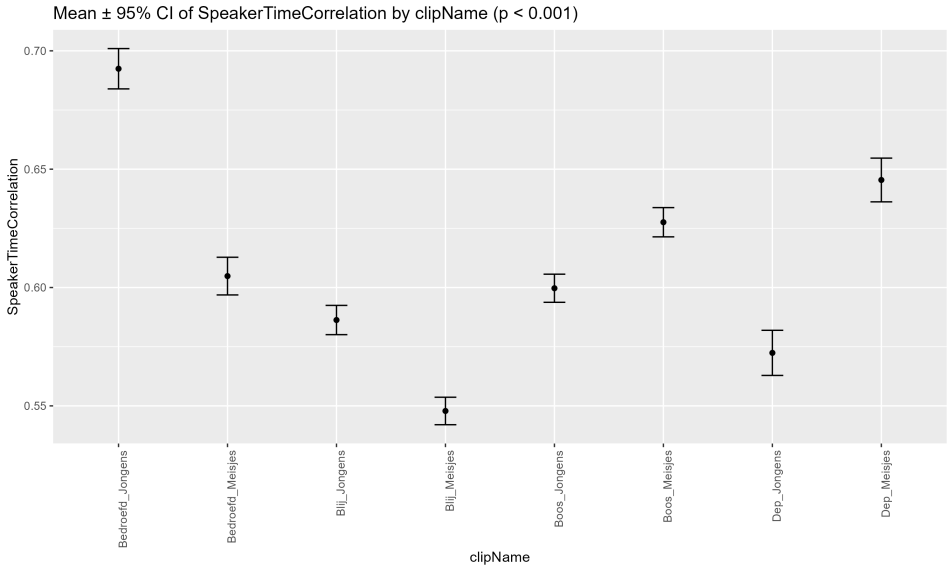


Figure C.14: Mean and 95% CI of Speaker Time Correlation across video clips. This plot shows the extent to which participants' eye movements were temporally aligned with the speaking actor. Variability across clips may reflect differences in how prominently or consistently the speaker was featured, influencing viewers' attention and engagement with speech-related visual cues.

Time Per Fixation

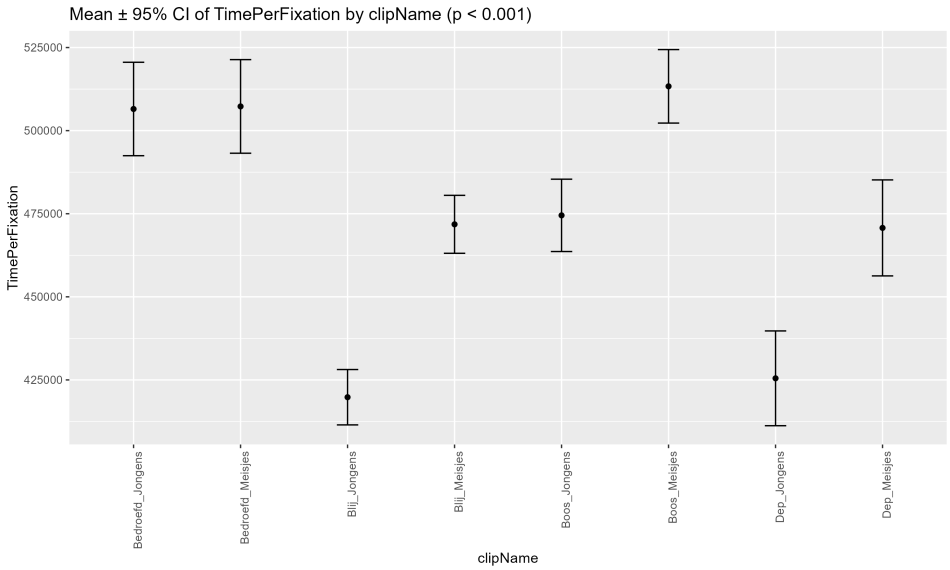


Figure C.15: Mean and 95% CI of Time Per Fixation across video clips. This feature captures the average duration that gaze remains on a single fixation point before moving on. Differences across clips may reflect variations in speaking pace, scene complexity, or viewer engagement.

D

Gender-Related Effects

Background Fixation

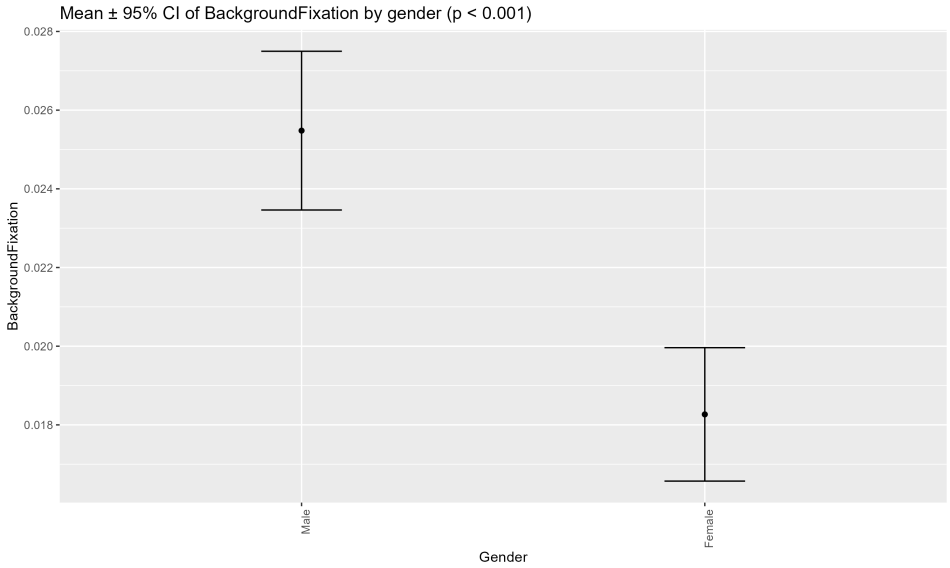


Figure D.1: Mean values with 95% CI per participant gender for the fixations on the background.

Body Fixation

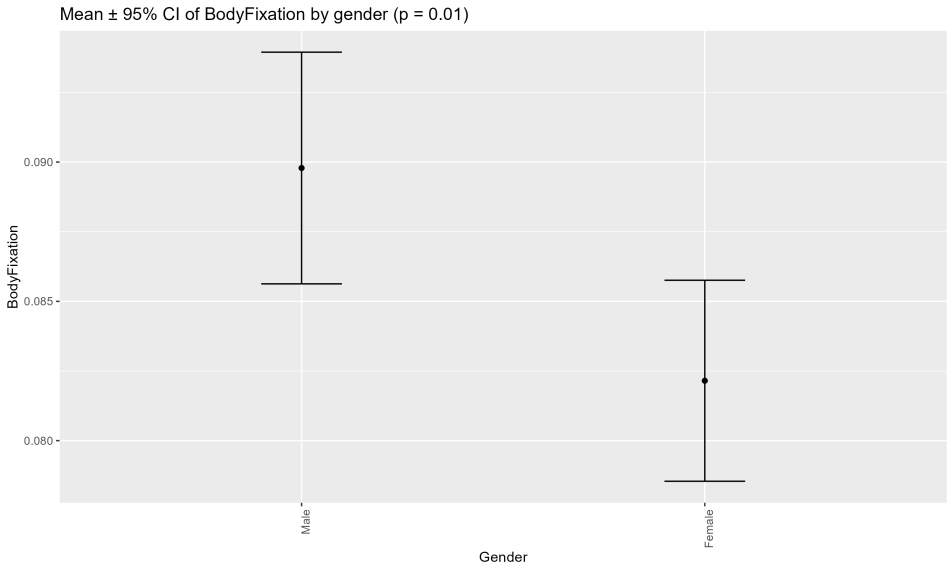


Figure D.2: Mean values with 95% CI per participant gender for the fixations on the body, excluding face, of the characters.

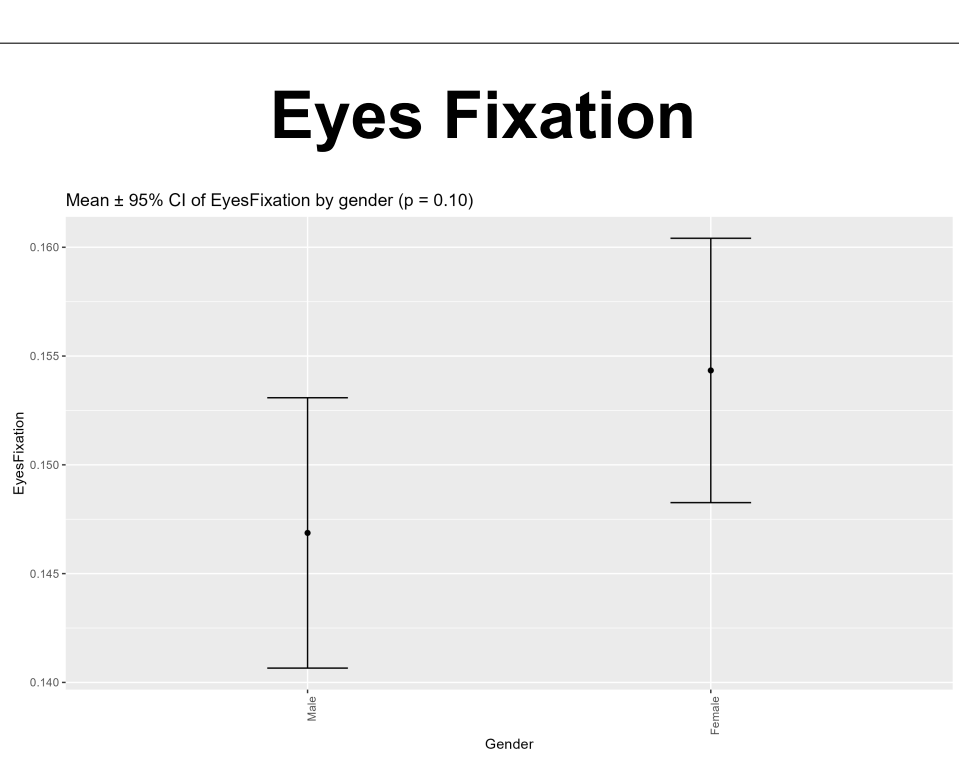


Figure D.3: Mean values with 95% confidence intervals (CI) per participant gender for the fixations on the eyes.

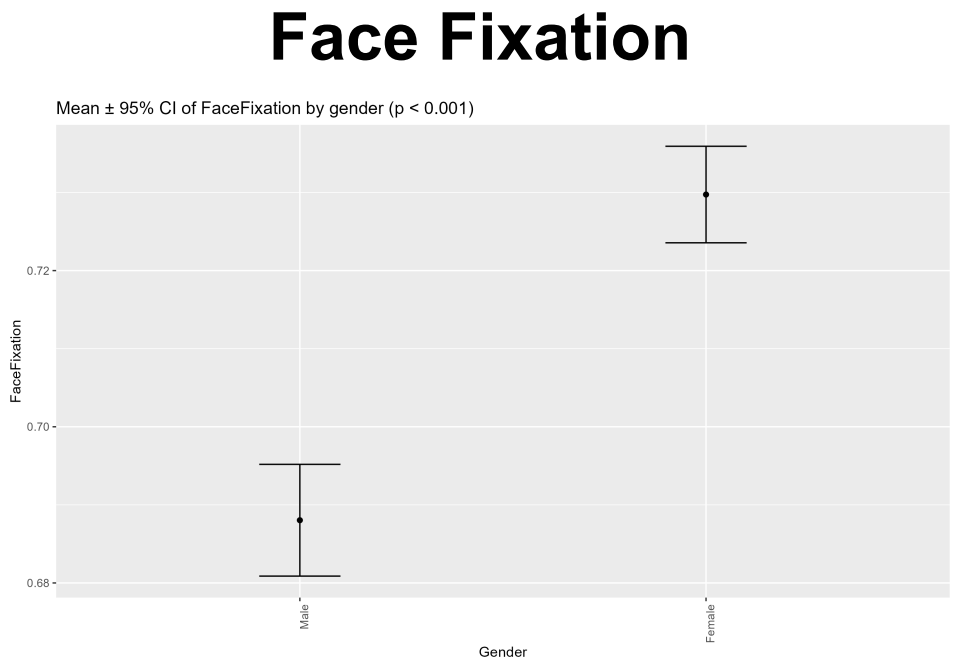


Figure D.4: Mean values with 95% CI per participant gender for the fixations on the characters faces.

Mouth Fixation

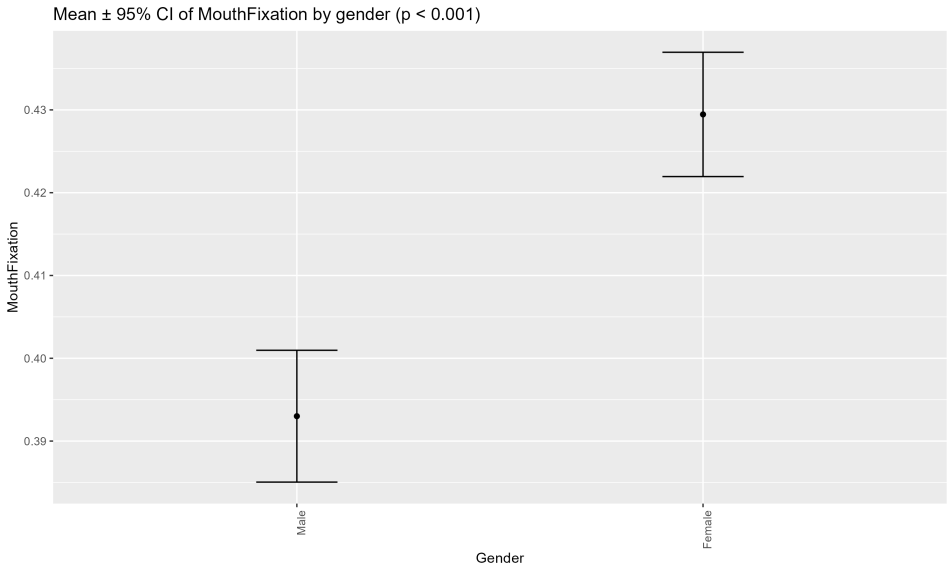


Figure D.5: Mean values with 95% CI per participant gender for the fixations on the characters mouths.

Non-Speaker Silhouette Correlation

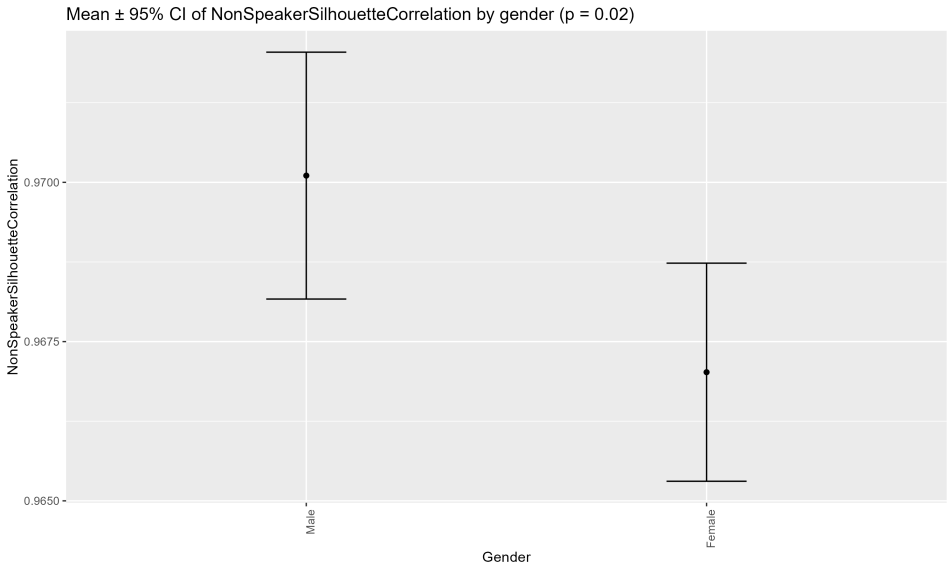


Figure D.6: Mean and 95% CI of Non-Speaker Silhouette Correlation across participant genders. This metric assesses how structurally aligned participants' gaze patterns were with the non-speaking individuals during speech, focusing on the overall shape of the gaze sequence rather than precise timing.

Non-Speaker Time Correlation

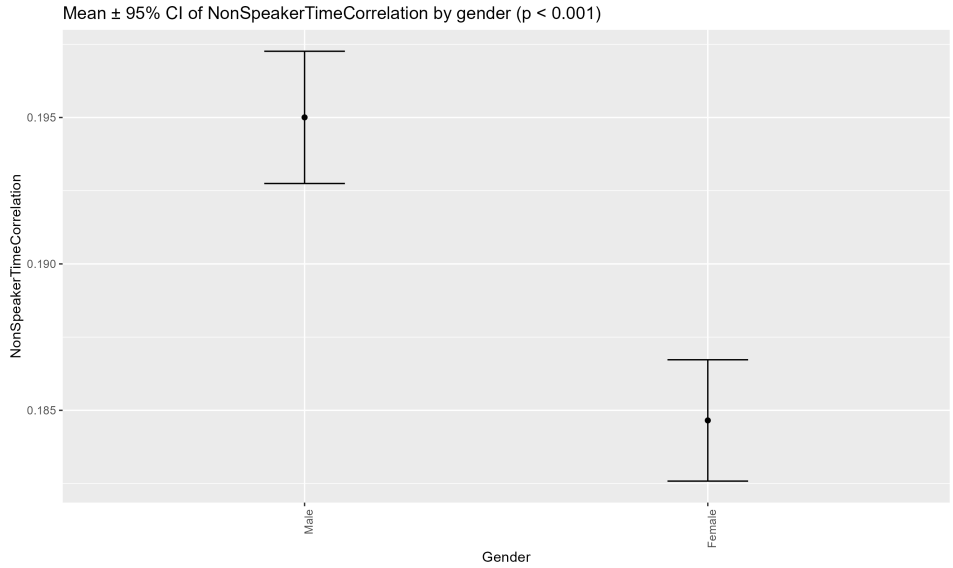


Figure D.7: Mean and 95% CI of Non-Speaker Time Correlation across participant genders. This figure illustrates the degree to which gaze behaviour tracked non-speaking individuals during speech.

Object Fixation

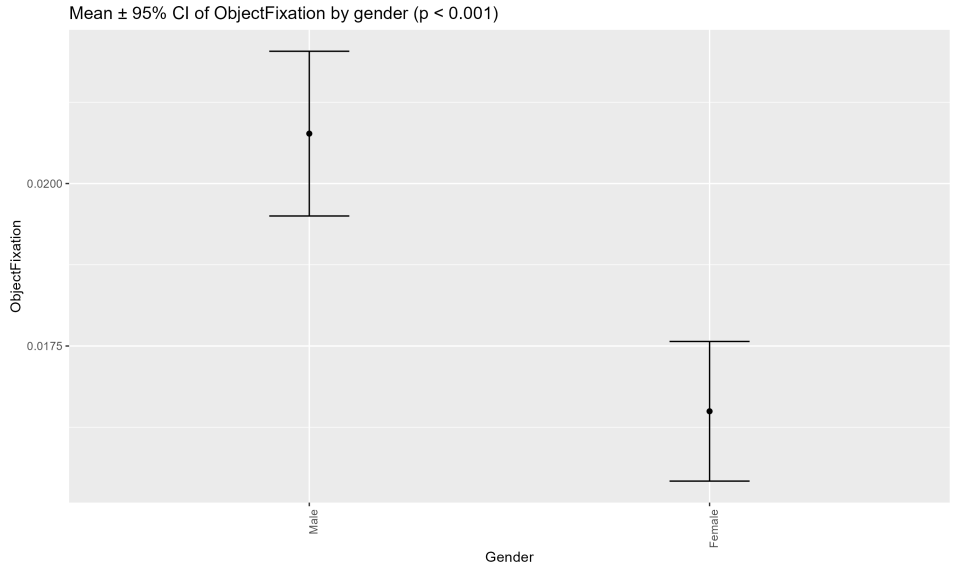


Figure D.8: Mean values with 95% CI per participant gender for the fixations on the distracting object in the scene.

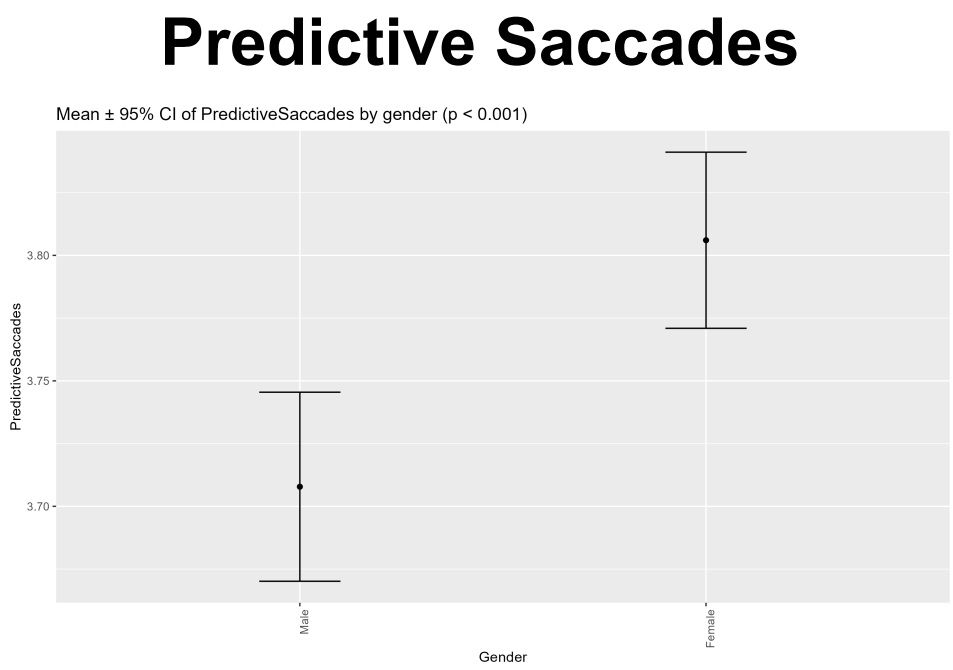


Figure D.9: Mean and 95% CI of Predictive Saccades across participant genders. Predictive saccades refer to anticipatory eye movements made ahead of speaker turns.

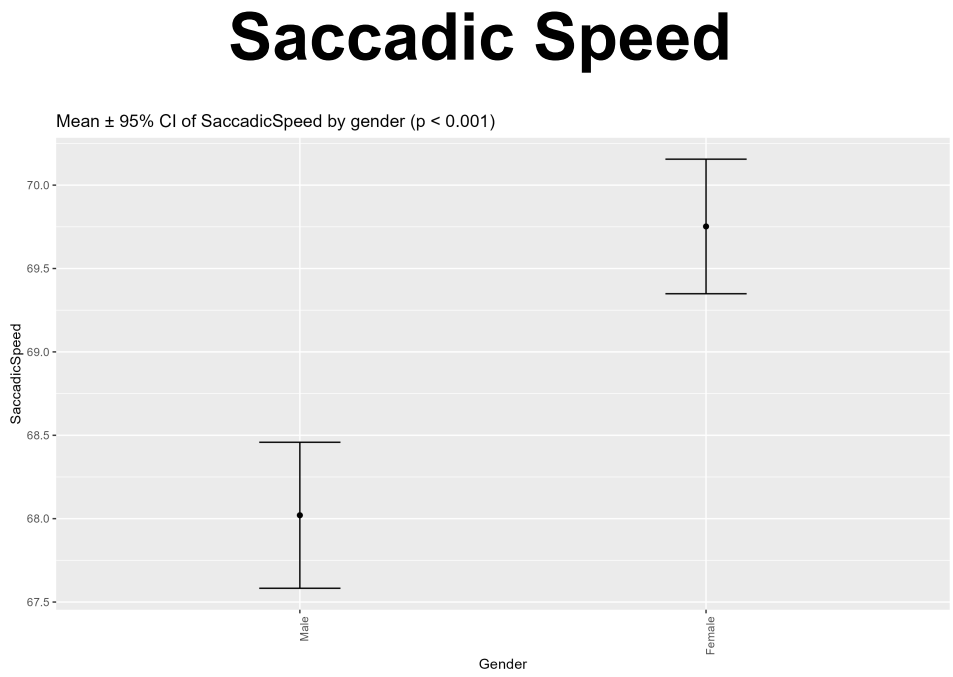


Figure D.10: Mean and 95% CI of Saccadic Speed across participant genders. Saccadic speed reflects the average velocity of rapid eye movements between fixations.

Scanpath Variance

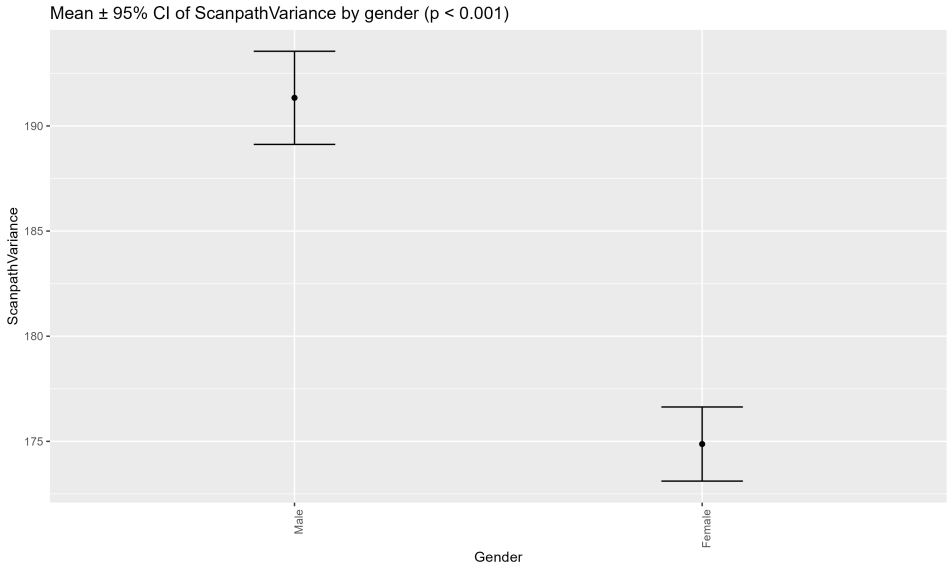


Figure D.11: Mean and 95% CI of Scanpath Variance across participant genders. This metric assesses how structurally aligned participants' gaze patterns were with the non-speaking individuals during speech, focusing on the overall shape of the gaze sequence rather than precise timing. This metric quantifies how much the gaze position of an individual deviates from the average across all participants.

Screen Time

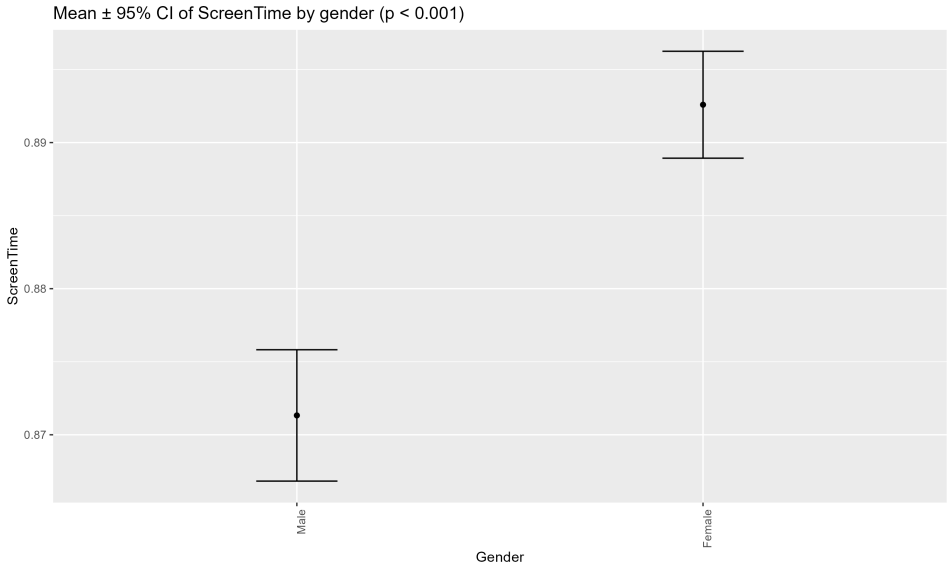


Figure D.12: Mean and 95% CI of Screen Time across participant genders. This plot displays the average proportion of time participants spent looking at the screen during each video clip, with error bars indicating the 95% CI.

Speaker Silhouette Correlation

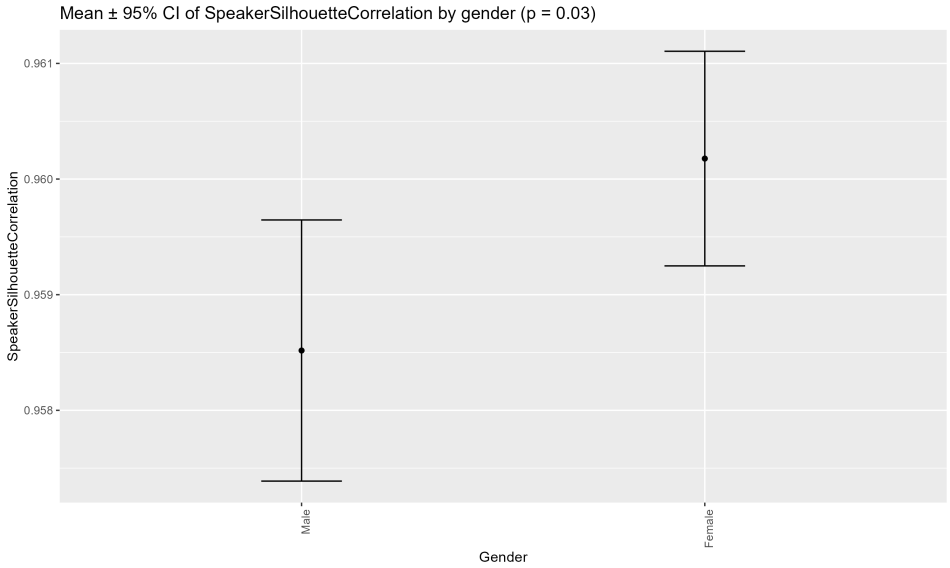


Figure D.13: Mean and 95% CI of Speaker Silhouette Correlation across participant genders. This measure captures the structural similarity between the participant's gaze pattern and the speaker's speech activity, reflecting the shape and continuity of engagement rather than just temporal overlap. Unlike time correlation, silhouette correlation penalises fragmented viewing behaviour, such as brief glances.

Speaker Time Correlation

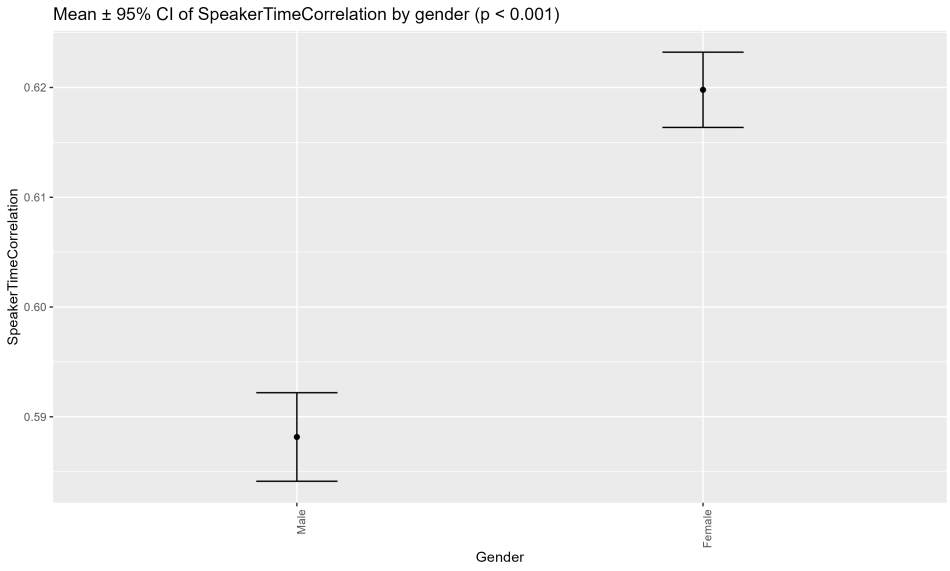


Figure D.14: Mean and 95% CI of Speaker Time Correlation across participant genders. This feature quantifies the extent to which participants' eye movements were temporally aligned with the speaking actor.

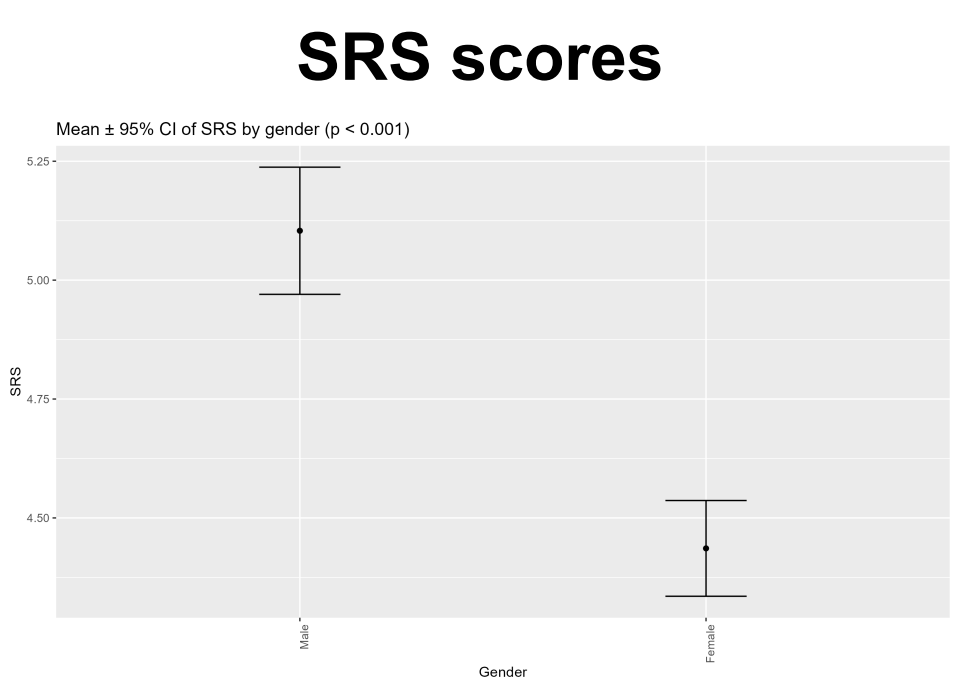


Figure D.15: Mean values of SRS scores with 95% CI per participant gender.

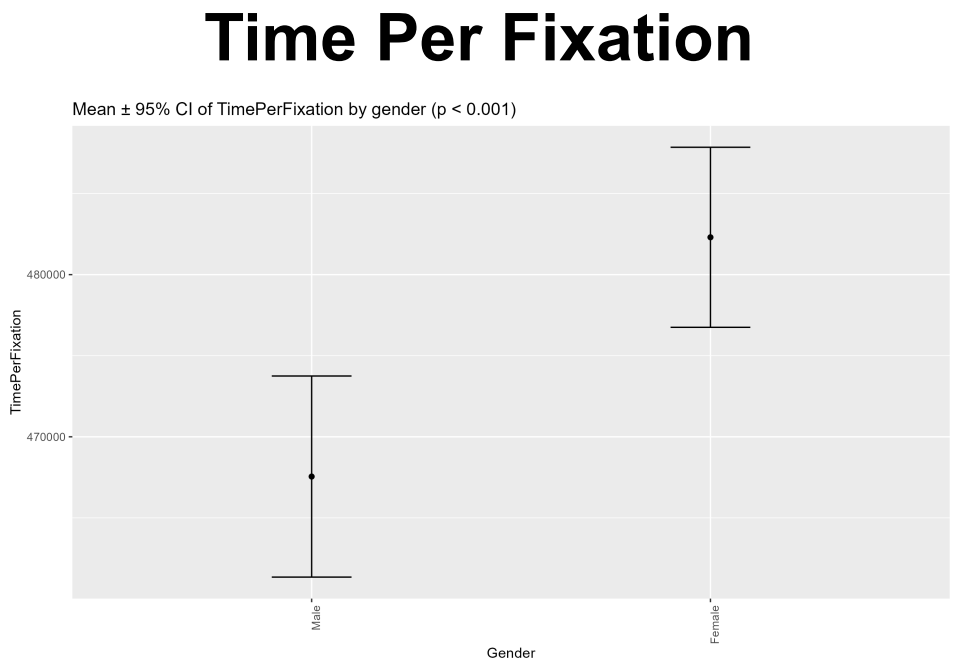


Figure D.16: Mean and 95% CI of Time Per Fixation across participant genders. This feature captures the average duration that gaze remains on a single fixation point before moving on.

E

Age-Related Effects

Background Fixation

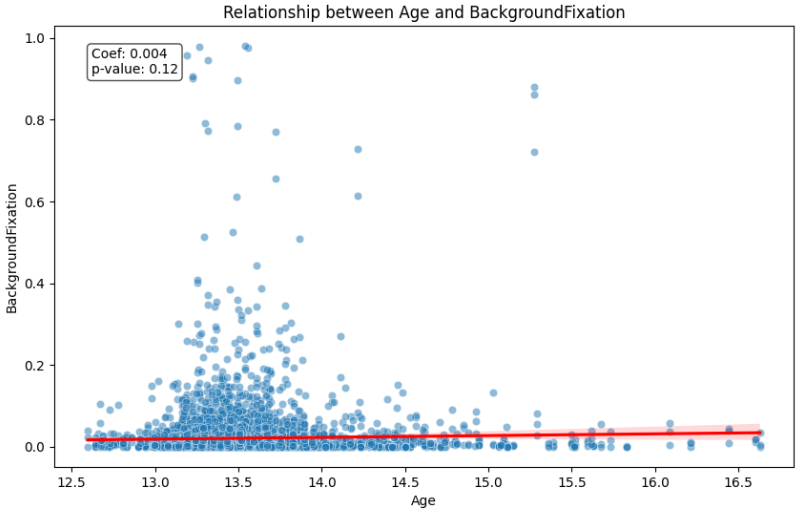


Figure E.1: Scatter plot with linear regression line showing the relationship between age and Background Fixation. Each point represents an individual measurement. The red line indicates the linear trend, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

Body Fixation

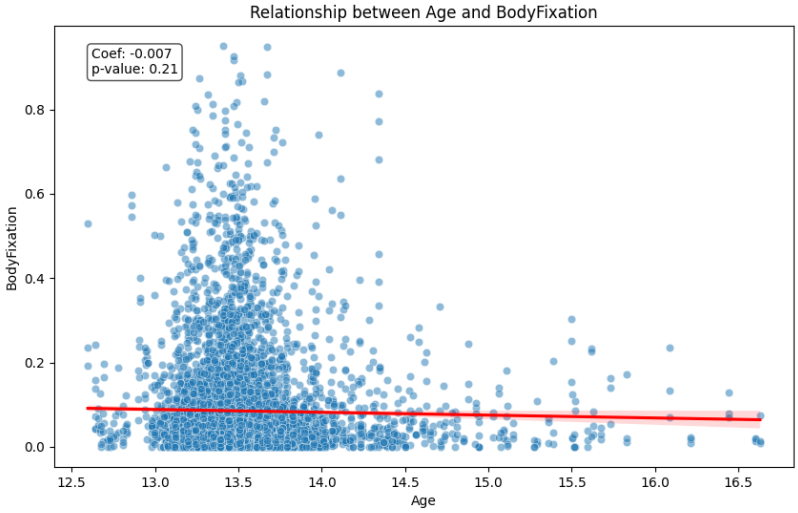


Figure E.2: Scatter plot with linear regression line showing the relationship between age and Body Fixation. Each point represents an individual measurement. The red line indicates the linear trend, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

Eyes Fixation

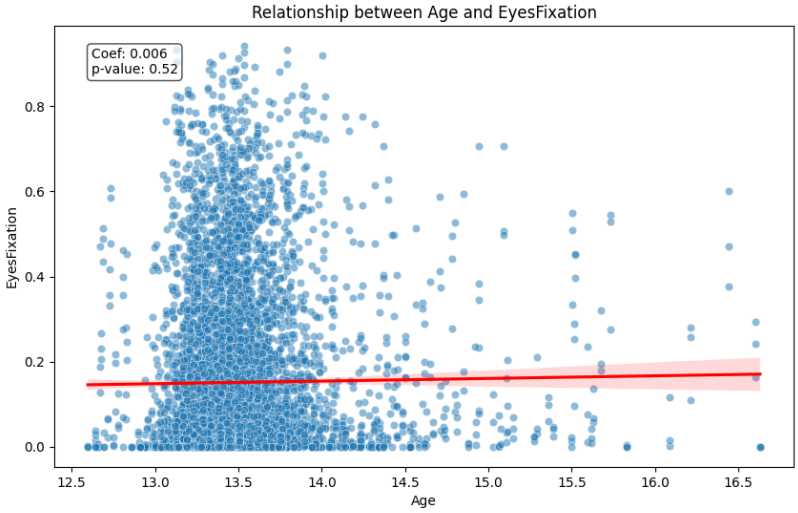


Figure E.3: Scatter plot with linear regression line showing the relationship between age and Eyes Fixation. Each point represents an individual measurement. The red line indicates the linear trend, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

Face Fixation

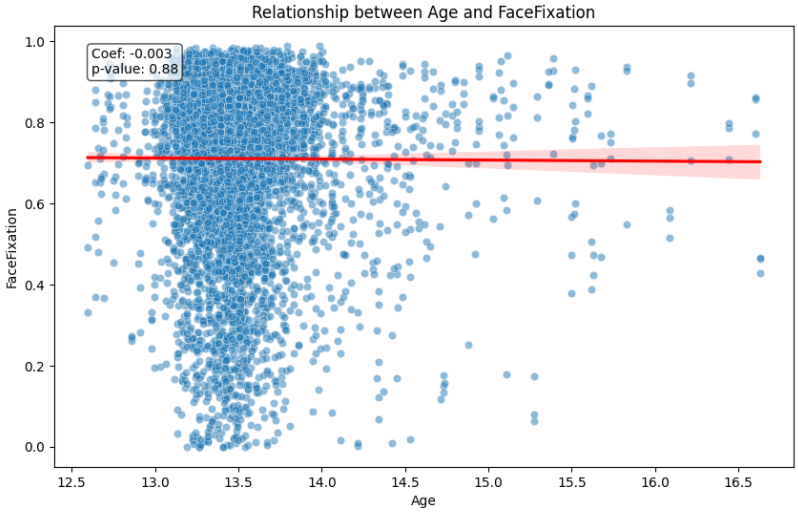


Figure E.4: Scatter plot with linear regression line showing the relationship between age and Face Fixation. Each point represents an individual measurement. The red line indicates the linear trend, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

Mouth Fixation

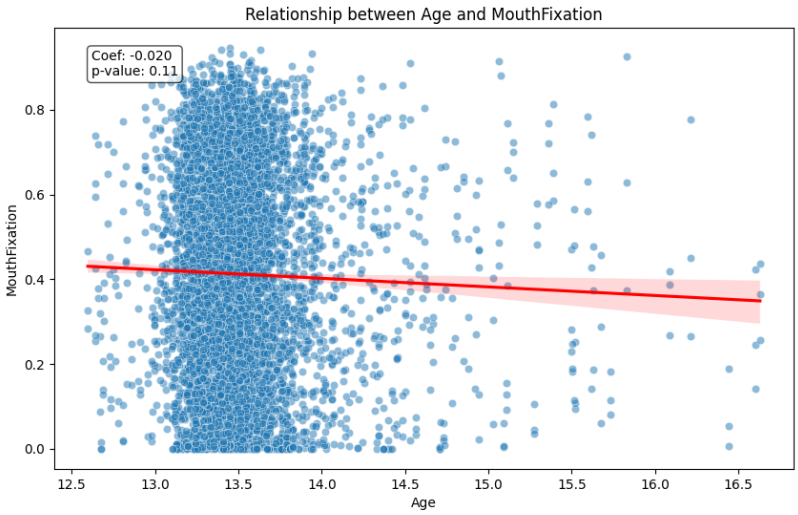


Figure E.5: Scatter plot with linear regression line showing the relationship between age and Mouth Fixation. Each point represents an individual measurement. The red line indicates the linear trend, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

Non-Speaker Silhouette Correlation

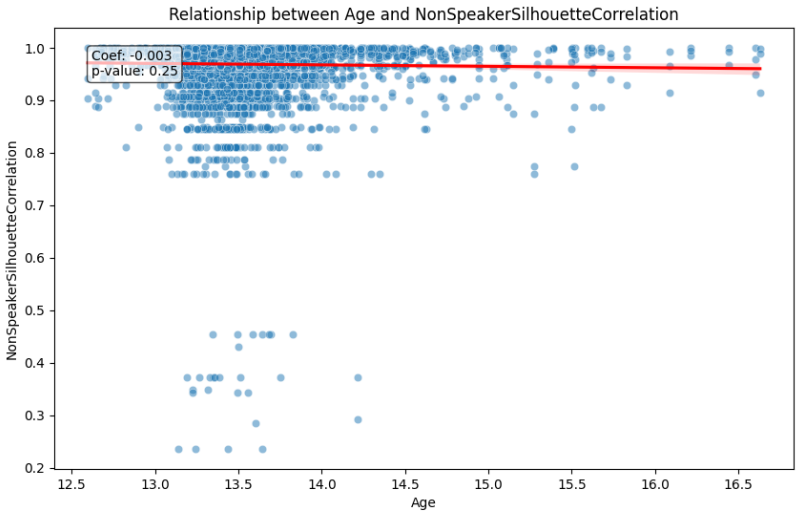


Figure E.6: Scatter plot with linear regression line showing the relationship between age and Non-Speaker Silhouette Correlation. Each point represents an individual measurement. The red line indicates the linear trend, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

Non-Speaker Time Correlation

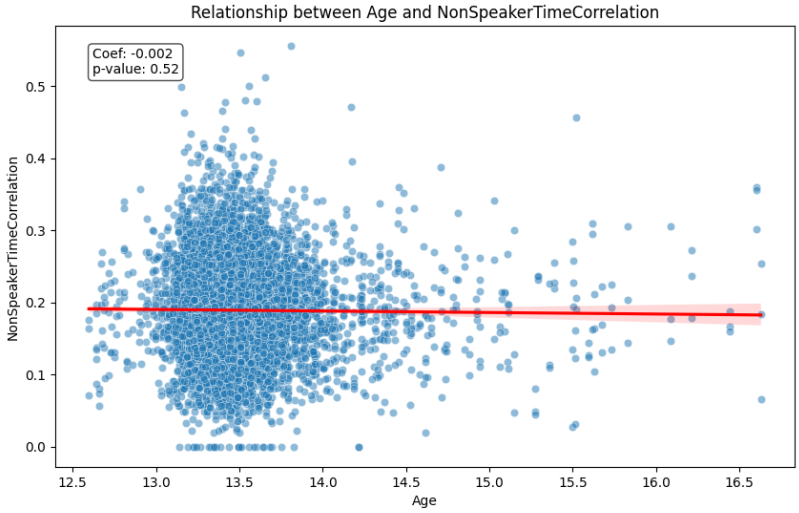


Figure E.7: Scatter plot with linear regression line showing the relationship between age and Non-Speaker Time Correlation. Each point represents an individual measurement. The red line indicates the linear trend, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

Object Fixation

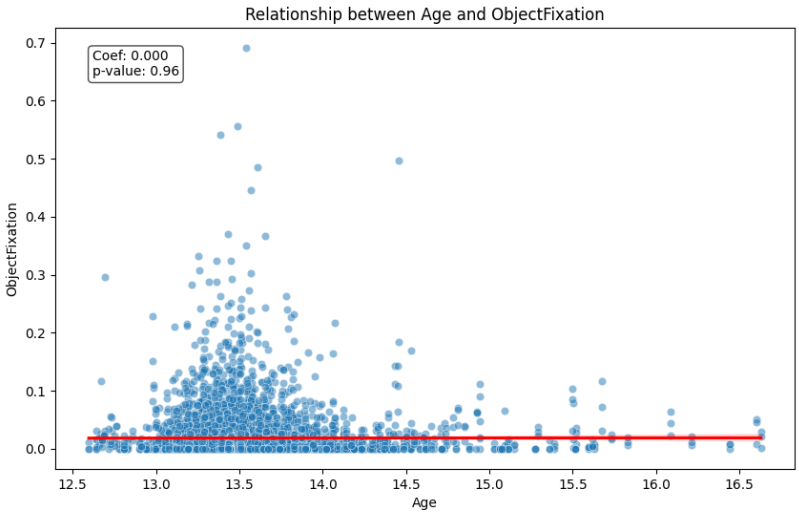


Figure E.8: Scatter plot with linear regression line showing the relationship between age and Object Fixation. Each point represents an individual measurement. The red line indicates the linear trend, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

Predictive Saccades

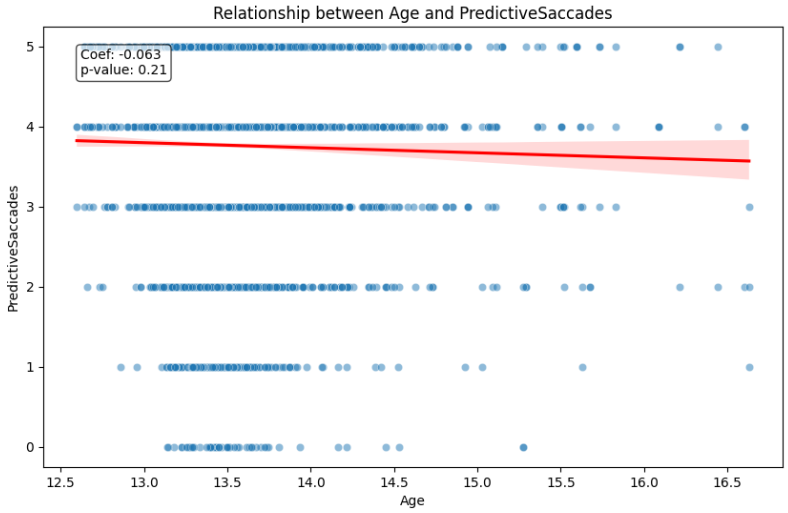


Figure E.9: Scatter plot with linear regression line showing the relationship between age and Predictive Saccades. Each point represents an individual measurement. The red line indicates the linear trend, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

Saccadic Speed

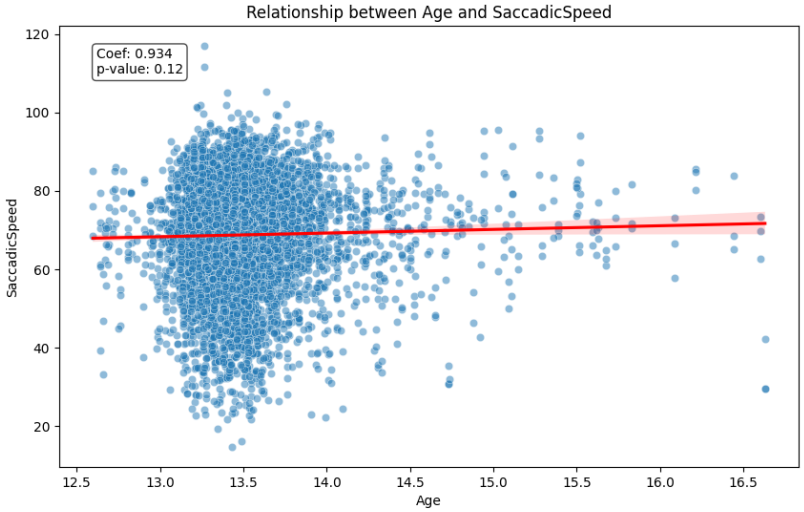


Figure E.10: Scatter plot with linear regression line showing the relationship between age and Saccadic Speed. Each point represents an individual measurement. The red line indicates the linear trend, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

Scanpath Variance

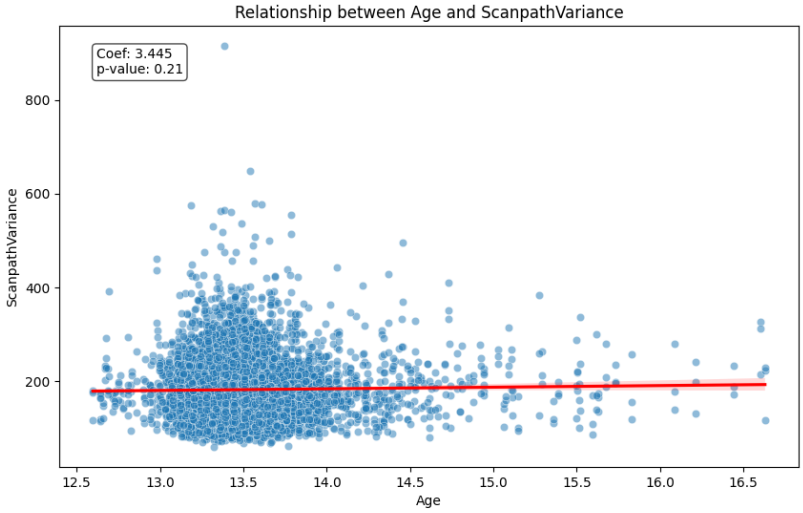


Figure E.11: Scatter plot with linear regression line showing the relationship between age and Scanpath Variance. Each point represents an individual measurement. The red line indicates the linear trend, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

Screen Time

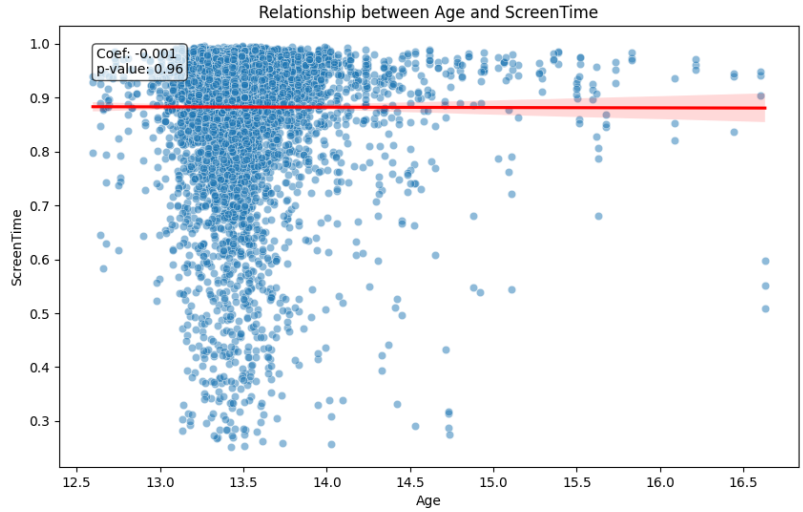


Figure E.12: Scatter plot with linear regression line showing the relationship between age and Screen Time. Each point represents an individual measurement. The red line indicates the linear trend, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

Speaker Silhouette Correlation

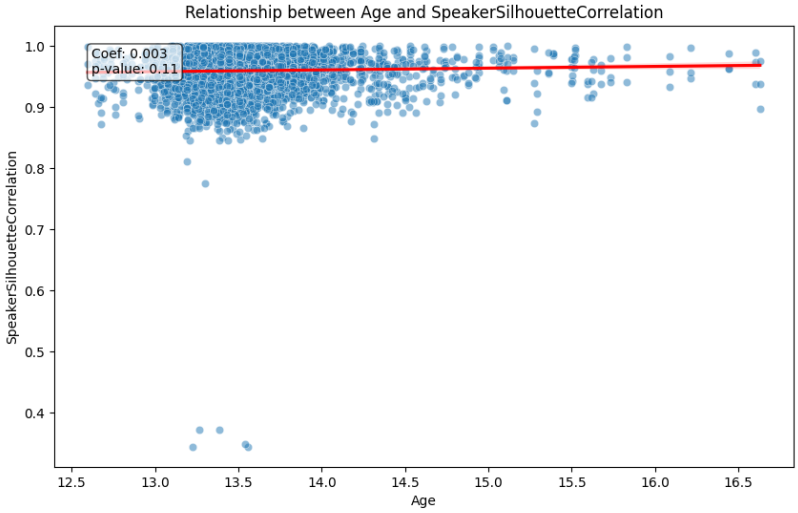


Figure E.13: Scatter plot with linear regression line showing the relationship between age and Speaker Silhouette Correlation. Each point represents an individual measurement. The red line indicates the linear trend, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

Speaker Time Correlation

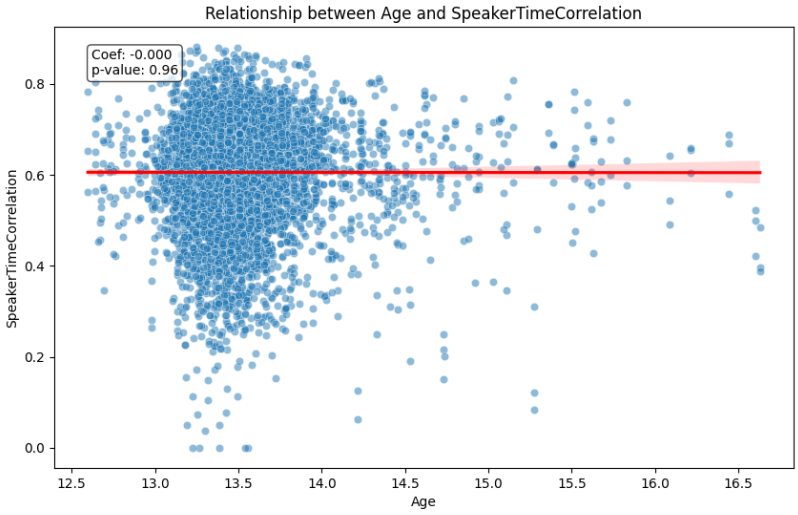


Figure E.14: Scatter plot with linear regression line showing the relationship between age and Speaker Time Correlation. Each point represents an individual measurement. The red line indicates the linear trend, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

SRS scores

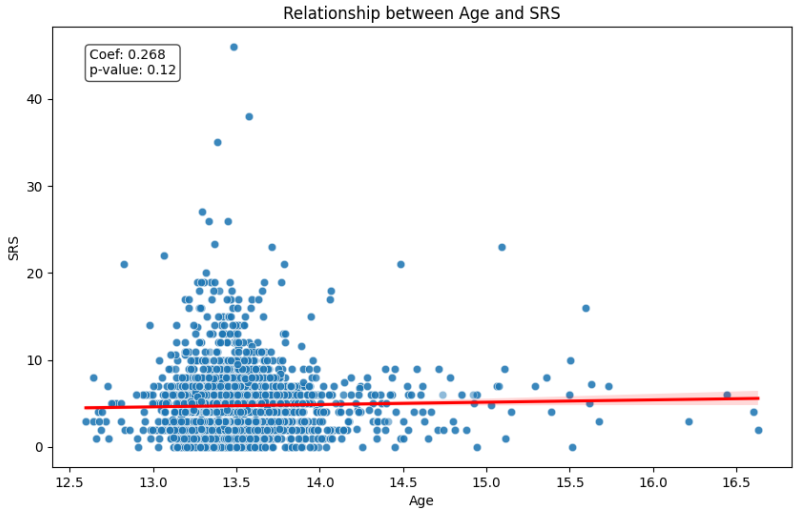


Figure E.15: Scatter plot with linear regression line showing the relationship between age and SRS scpres. Each point represents an individual measurement. The red line indicates the linear trend, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

Time Per Fixation

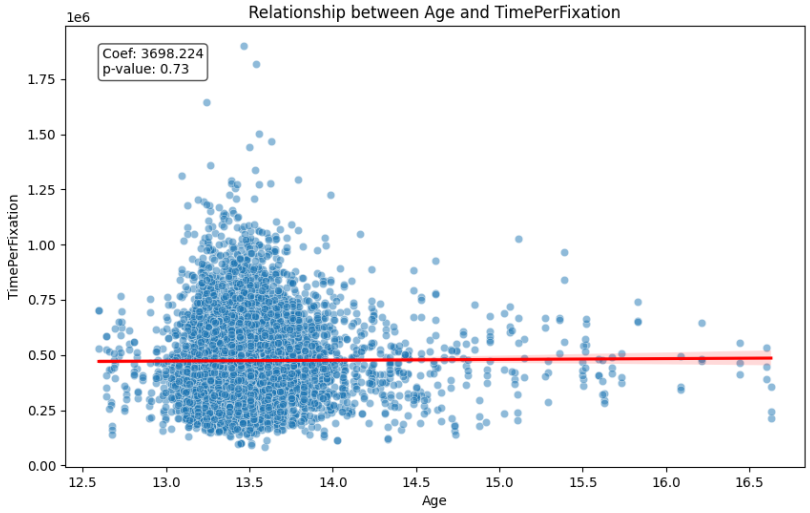


Figure E.16: Scatter plot with linear regression line showing the relationship between age and Time Per Fixation. Each point represents an individual measurement. The red line indicates the linear trend, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

F

GEE-GLM Autism Diagnosis: All Participants

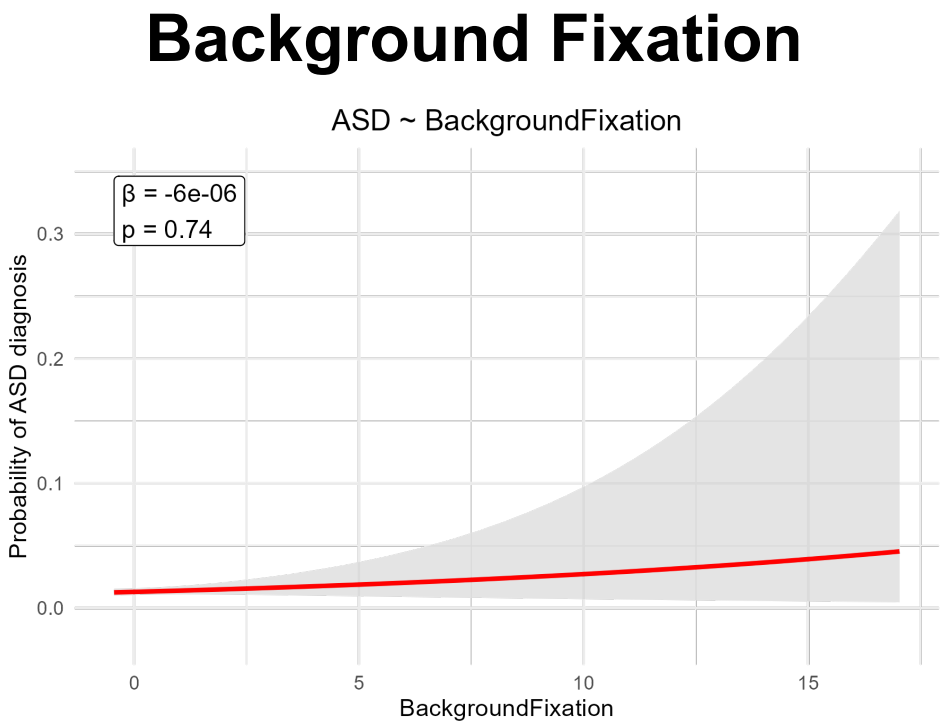


Figure F.1: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Background Fixation in all participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

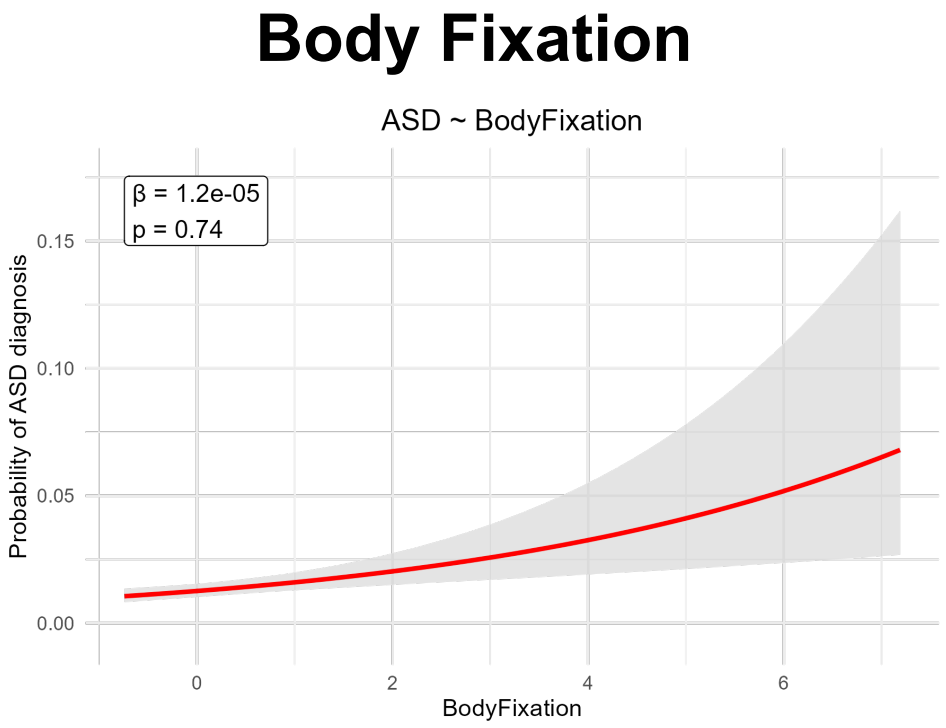


Figure F.2: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Body Fixation in all participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

Figure F.3: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Eyes Fixation in all participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

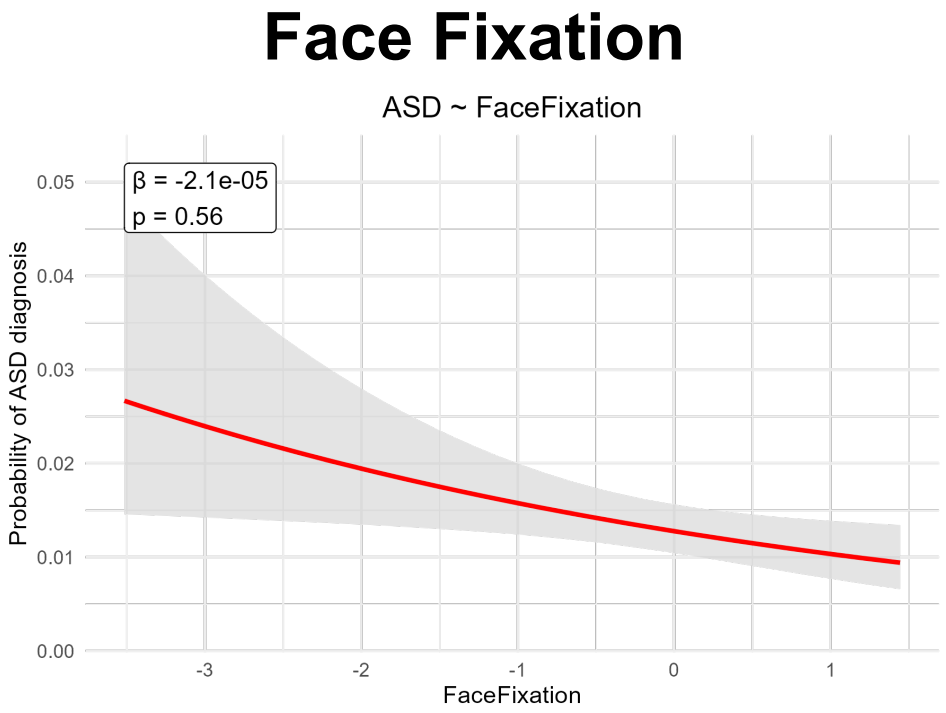


Figure F.4: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Face Fixation in all participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

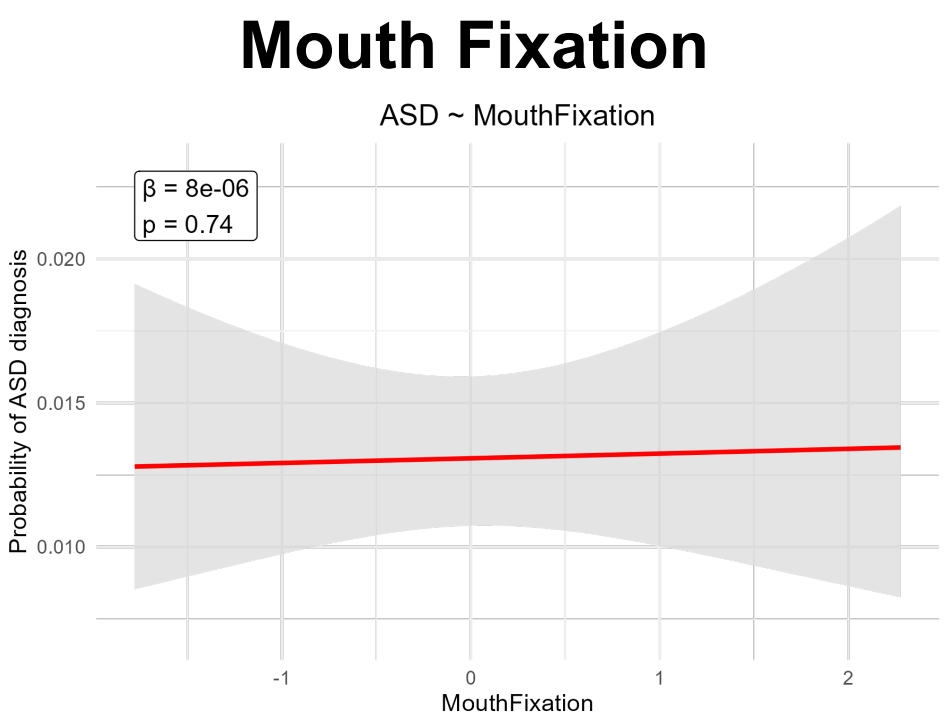


Figure F.5: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Mouth Fixation in all participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

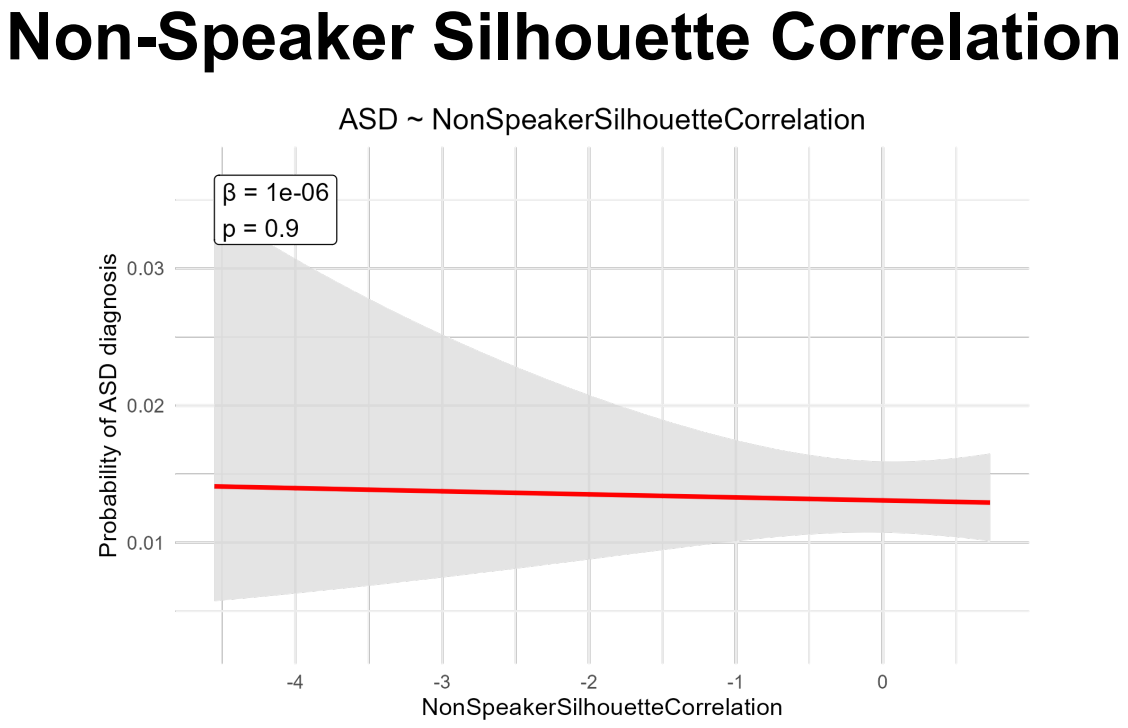


Figure F.6: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Non-Speaker Silhouette Correlation in all participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

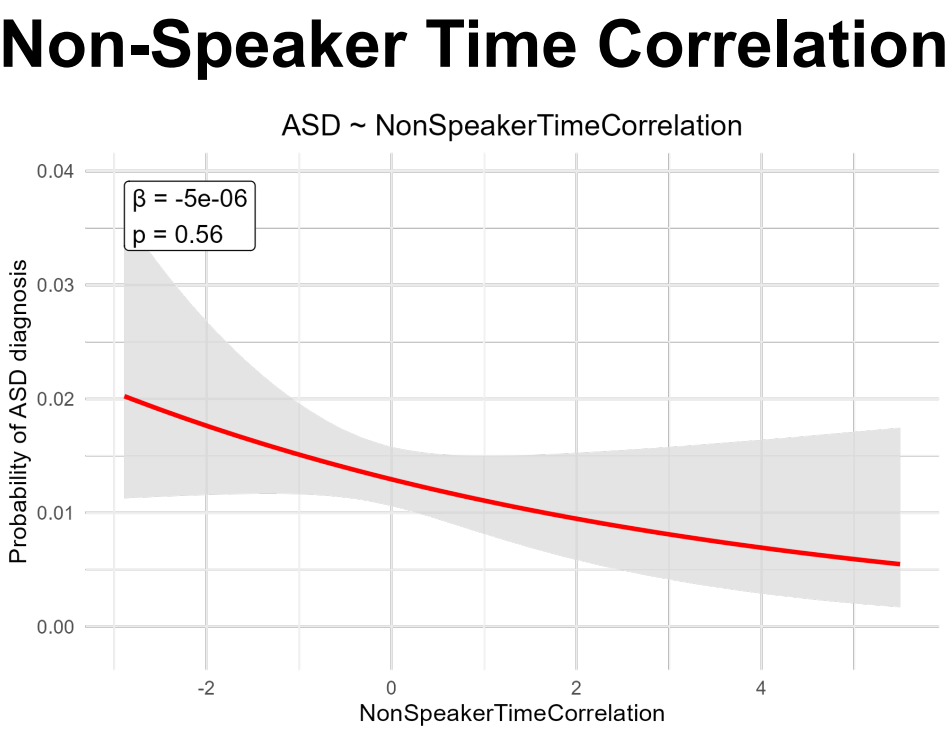


Figure F.7: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Non-Speaker Time Correlation in all participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

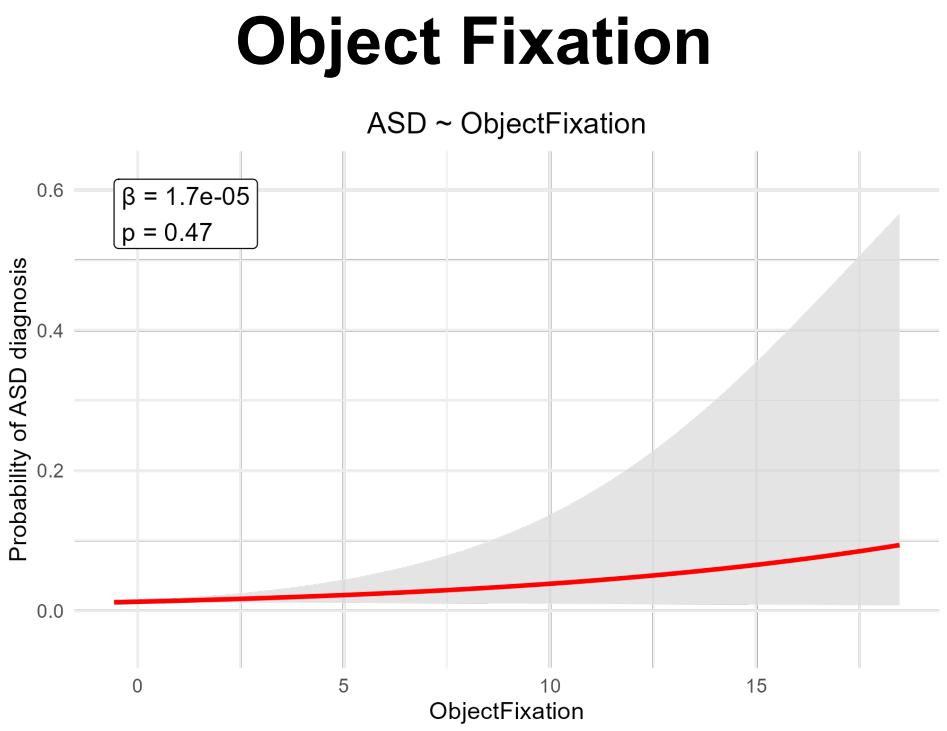


Figure F.8: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Object Fixation in all participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

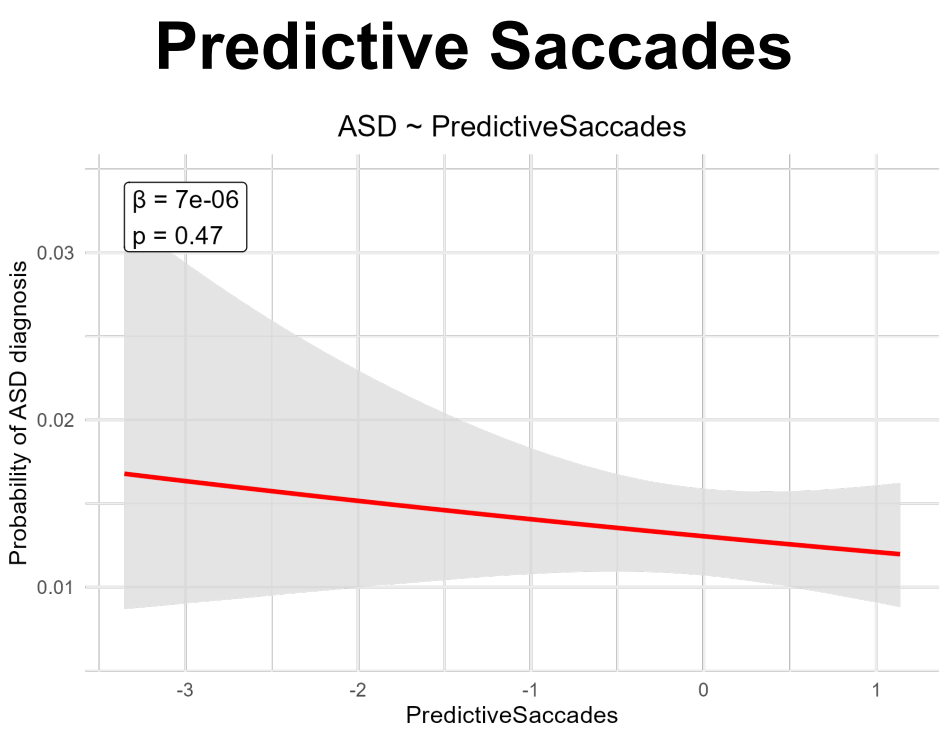


Figure F.9: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Predictive Saccades in all participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

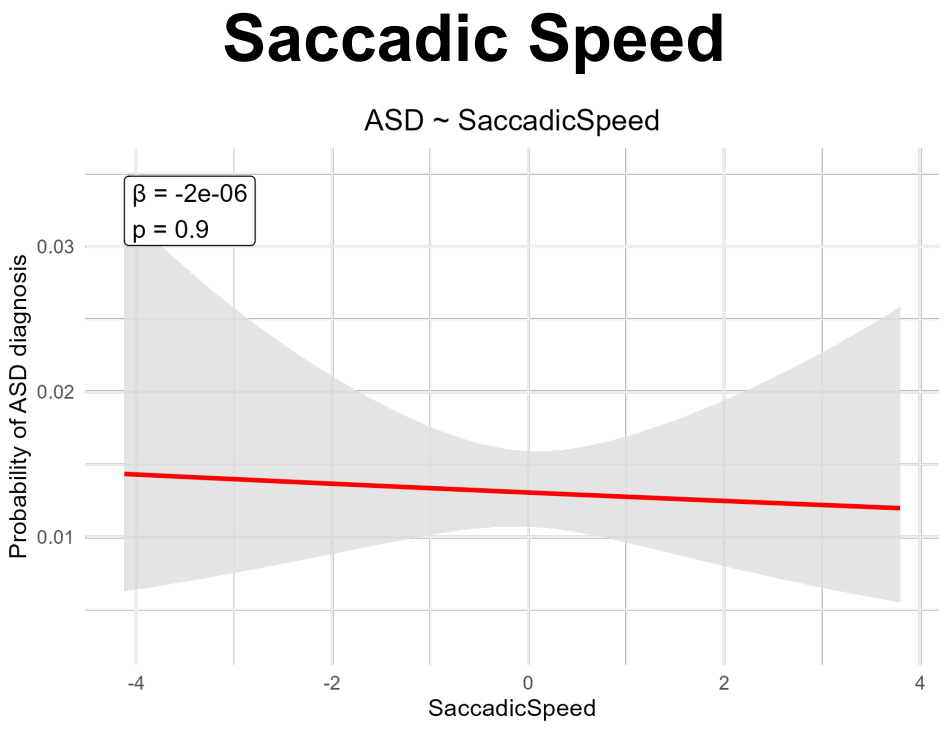


Figure F.10: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Saccadic Speed in all participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

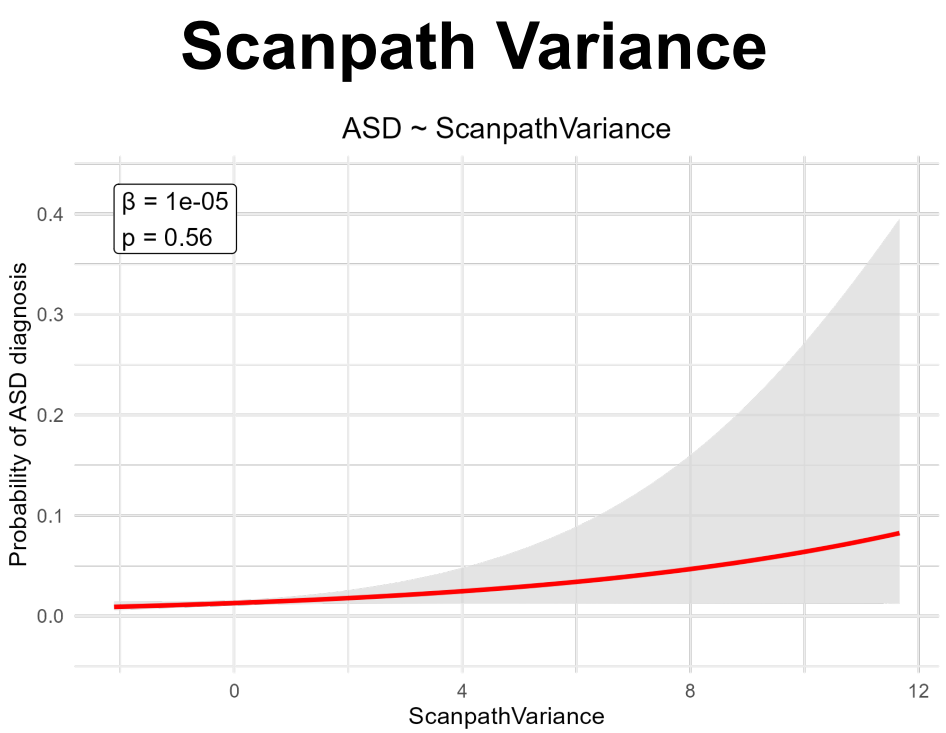


Figure F.11: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Scanpath Variance in all participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

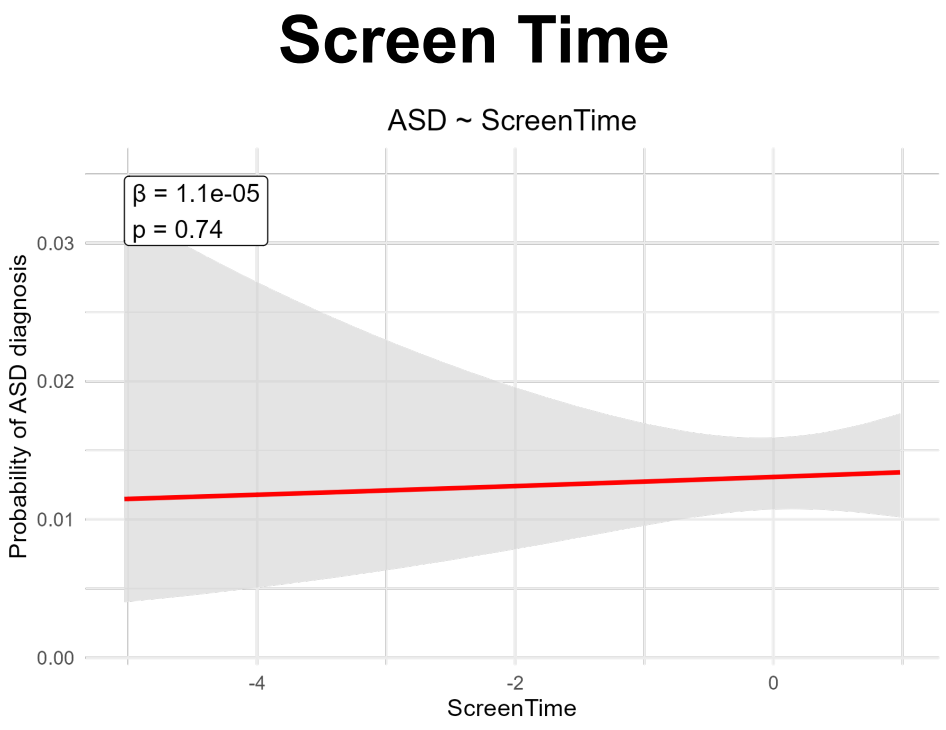


Figure F.12: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Screen Time in all participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

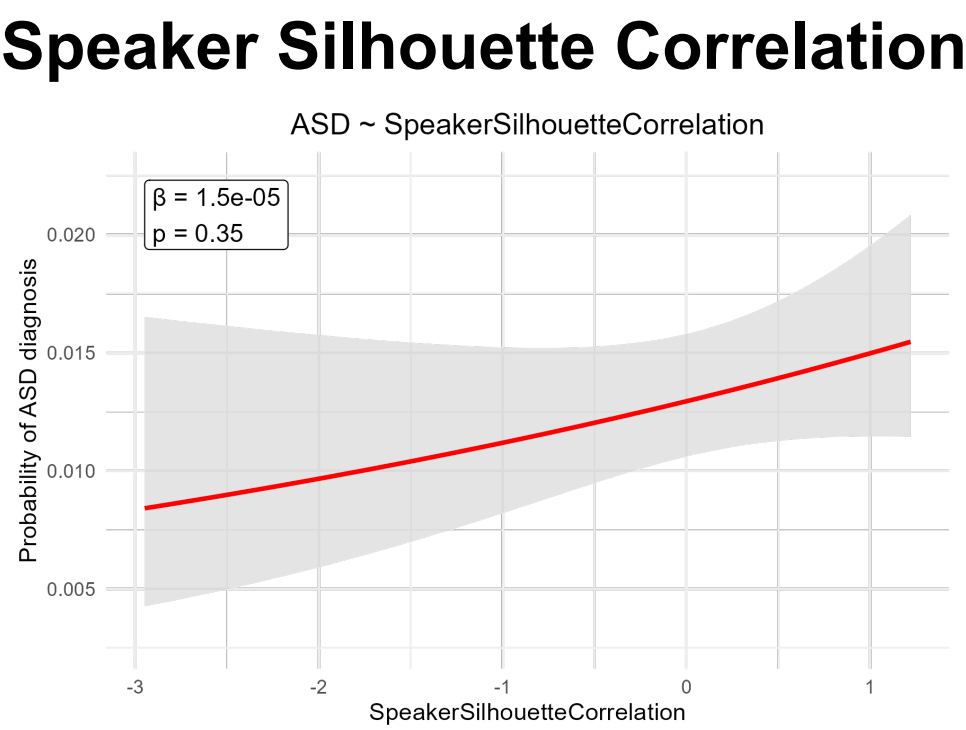


Figure F.13: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Speaker Silhouette Correlation in all participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

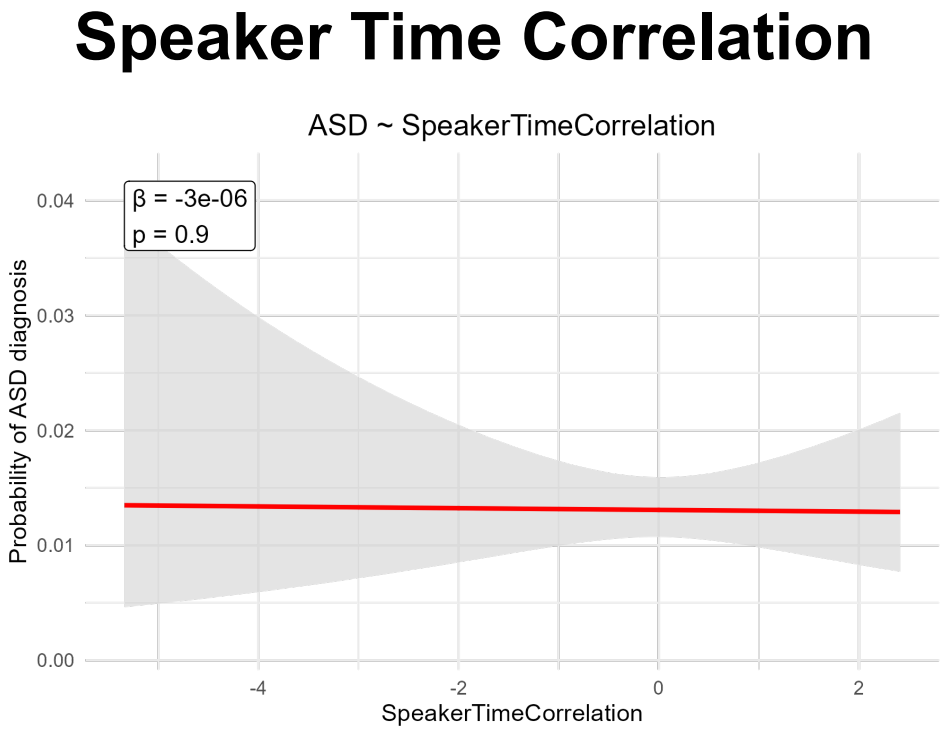


Figure F.14: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Speaker Time Correlation in all participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

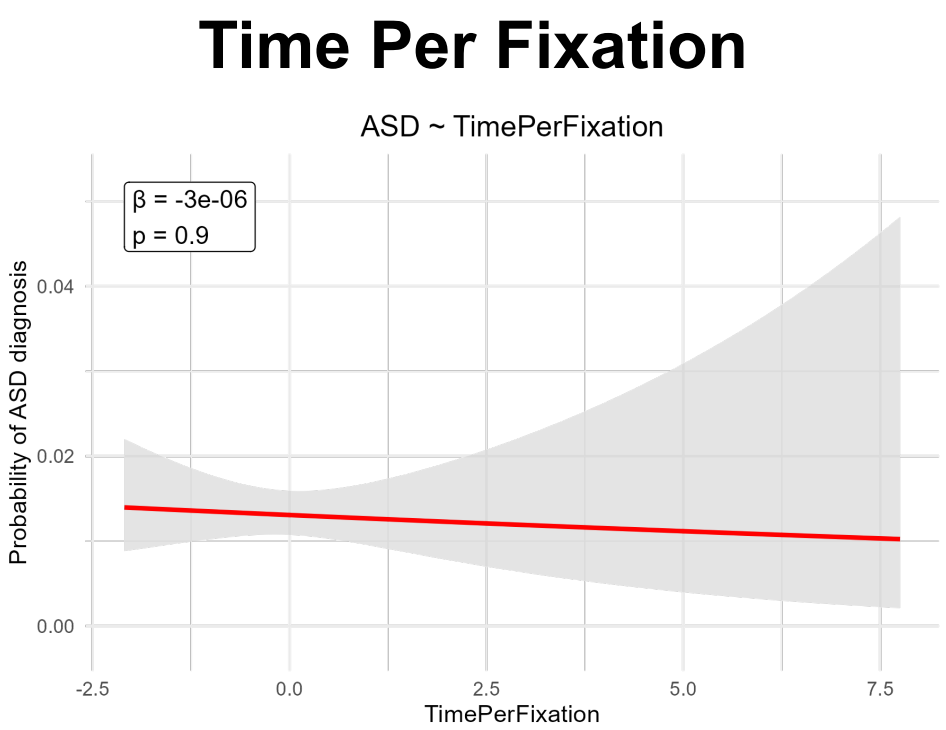


Figure F.15: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Time Per Fixation in all participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

G

GEE-GLM Autism Diagnosis: Male Only

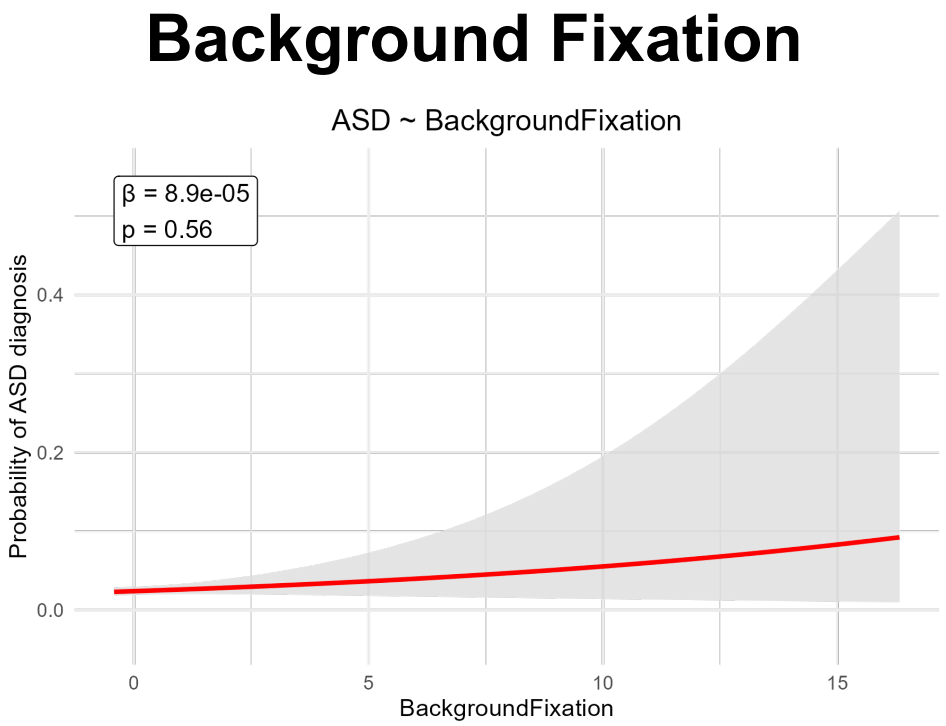


Figure G.1: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Background Fixation in male participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

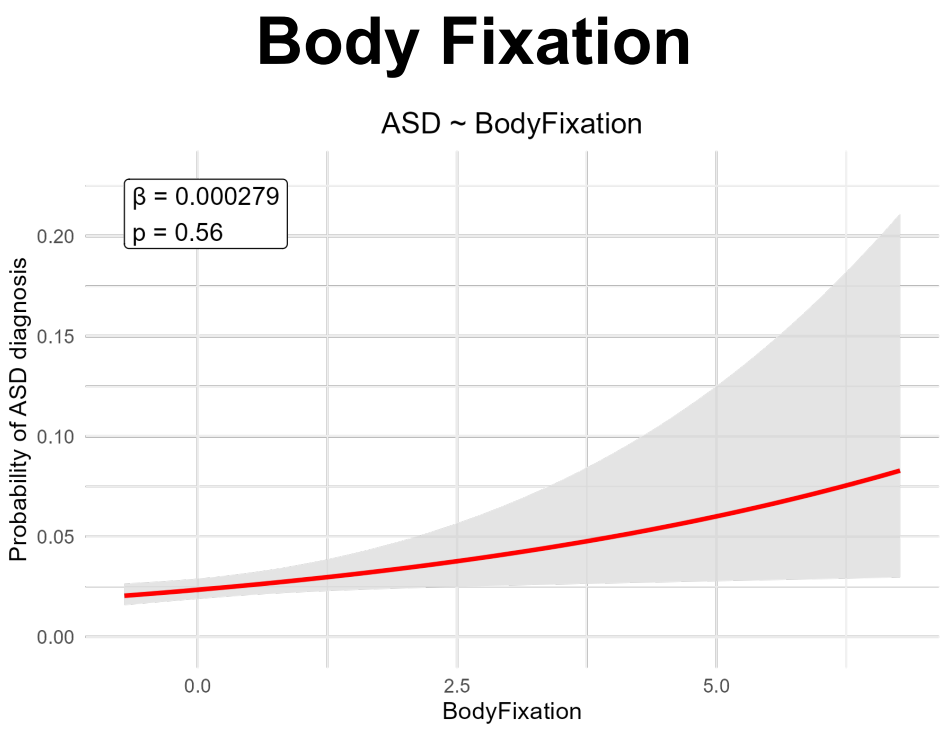


Figure G.2: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Body Fixation in male participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

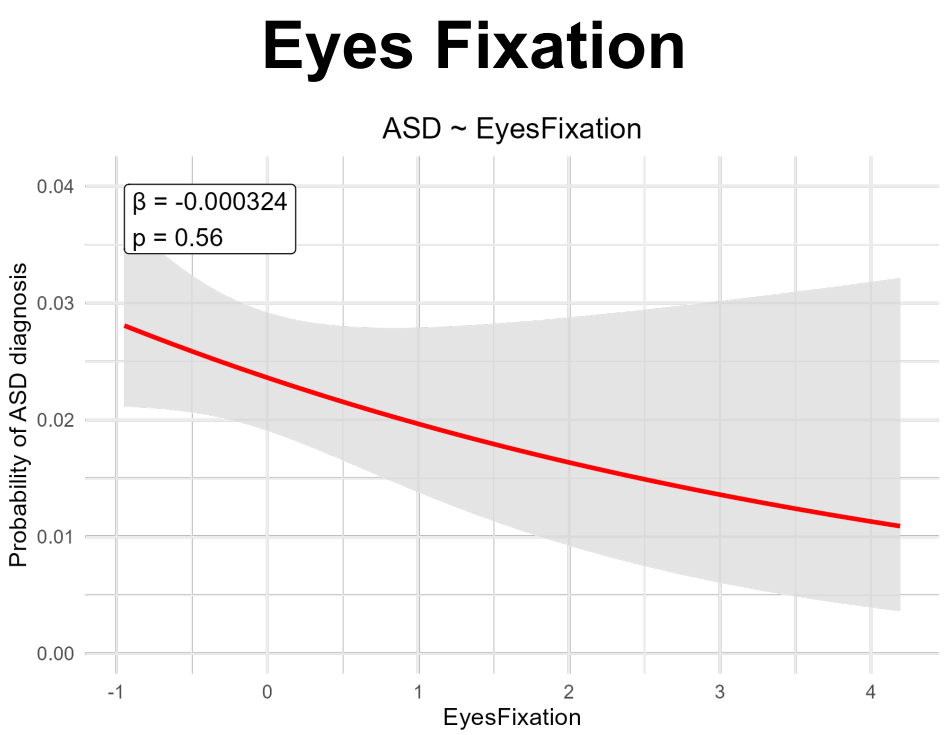


Figure G.3: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Eyes Fixation in male participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

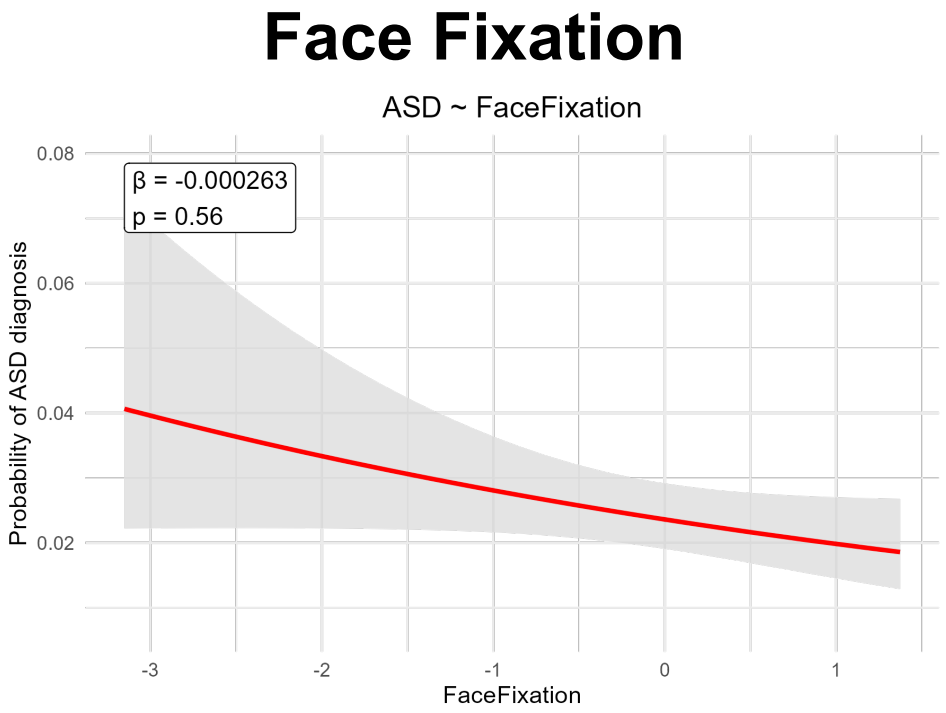


Figure G.4: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Face Fixation in male participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

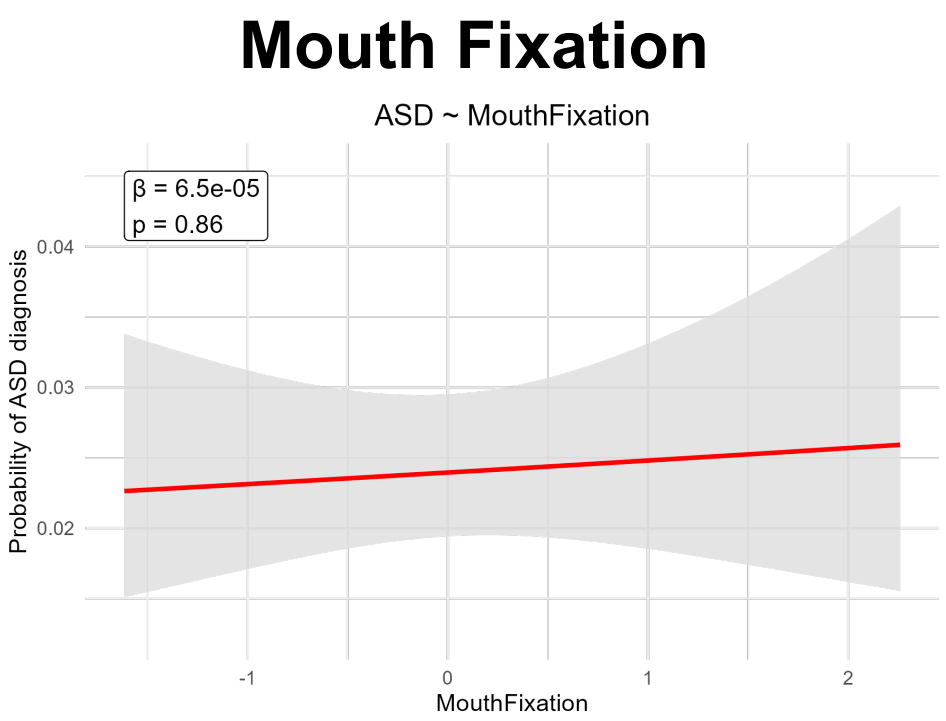


Figure G.5: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Mouth Fixation in male participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

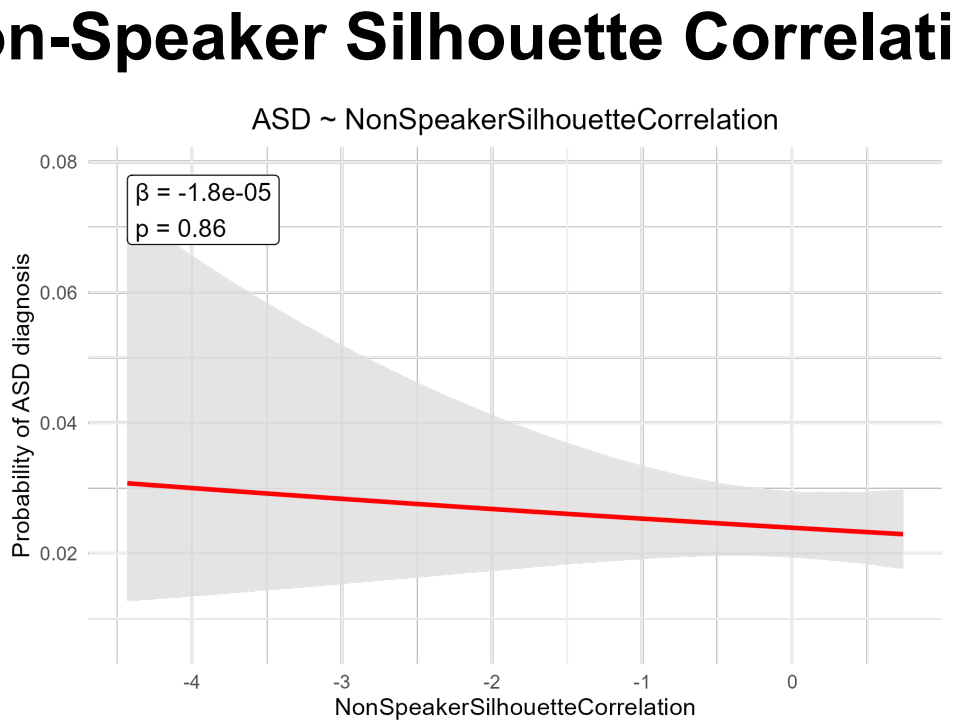


Figure G.6: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Non-Speaker Silhouette Correlation in male participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

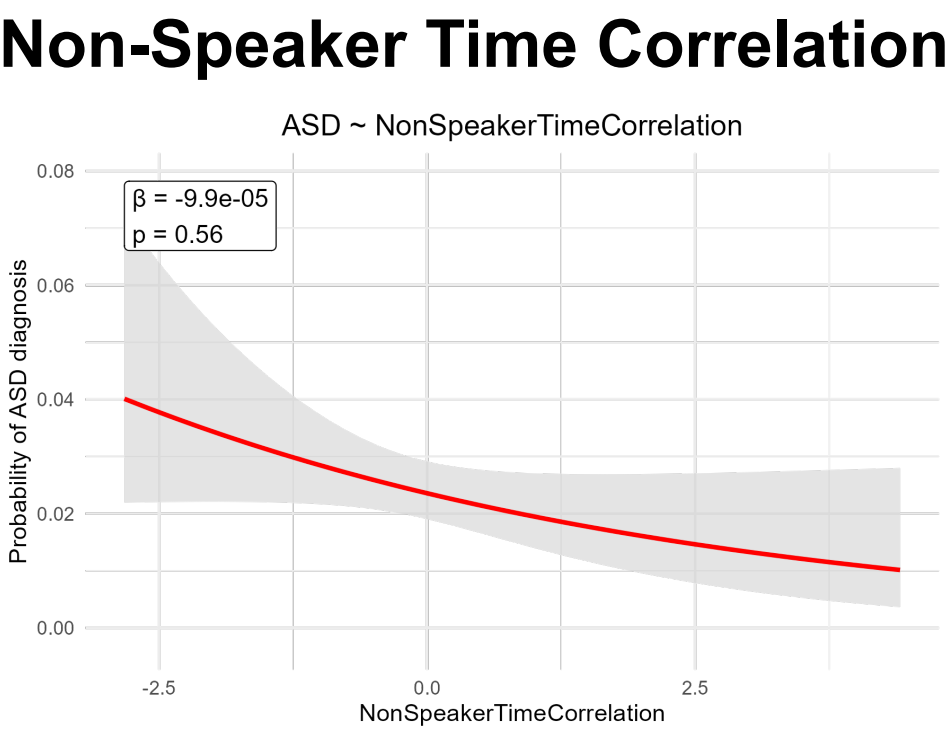


Figure G.7: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Non-Speaker Time Correlation in male participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

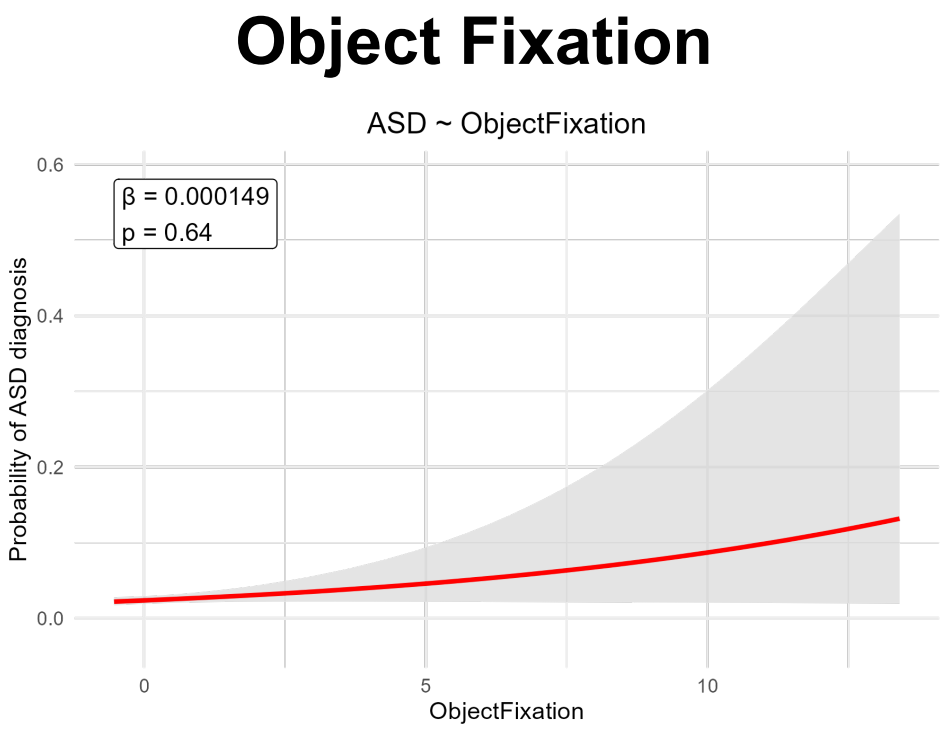


Figure G.8: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Object Fixation in male participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

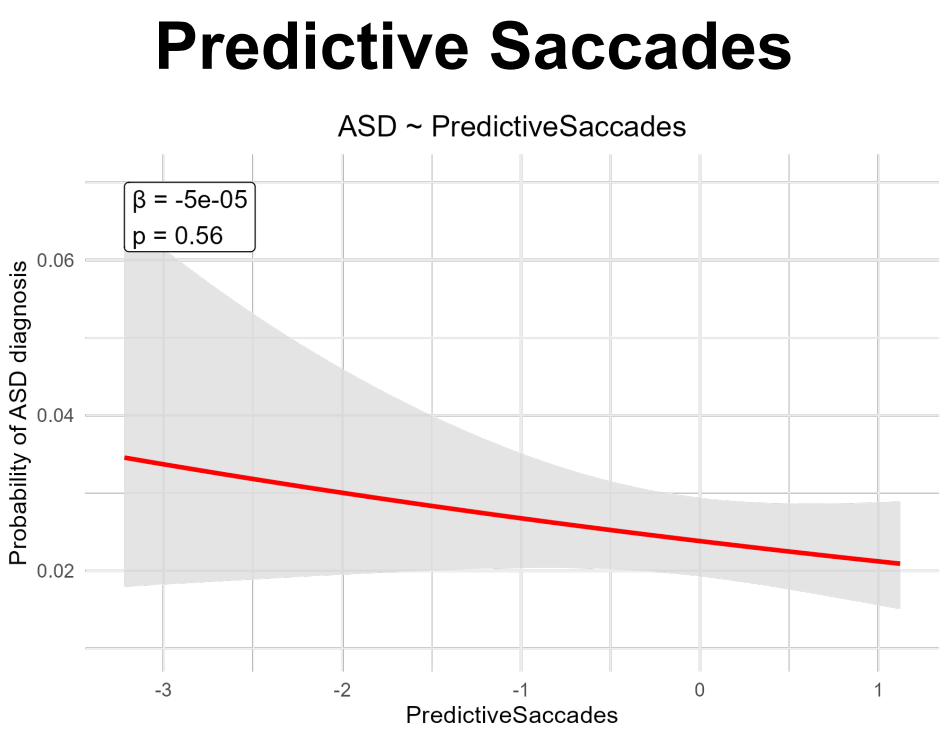


Figure G.9: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Predictive Saccades in male participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

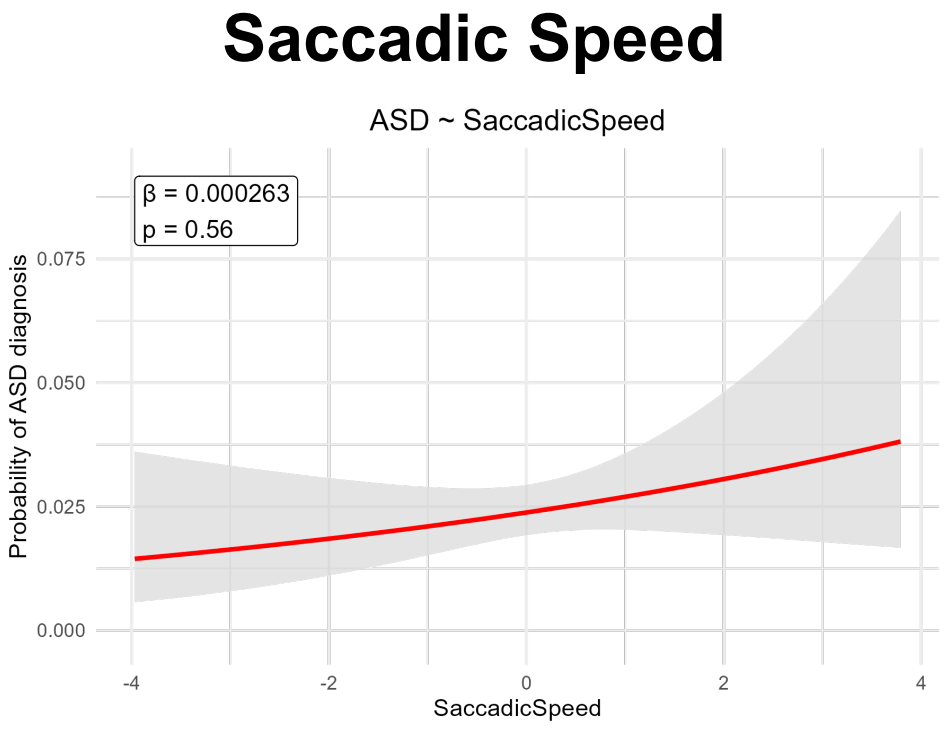


Figure G.10: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Saccadic Speed in male participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

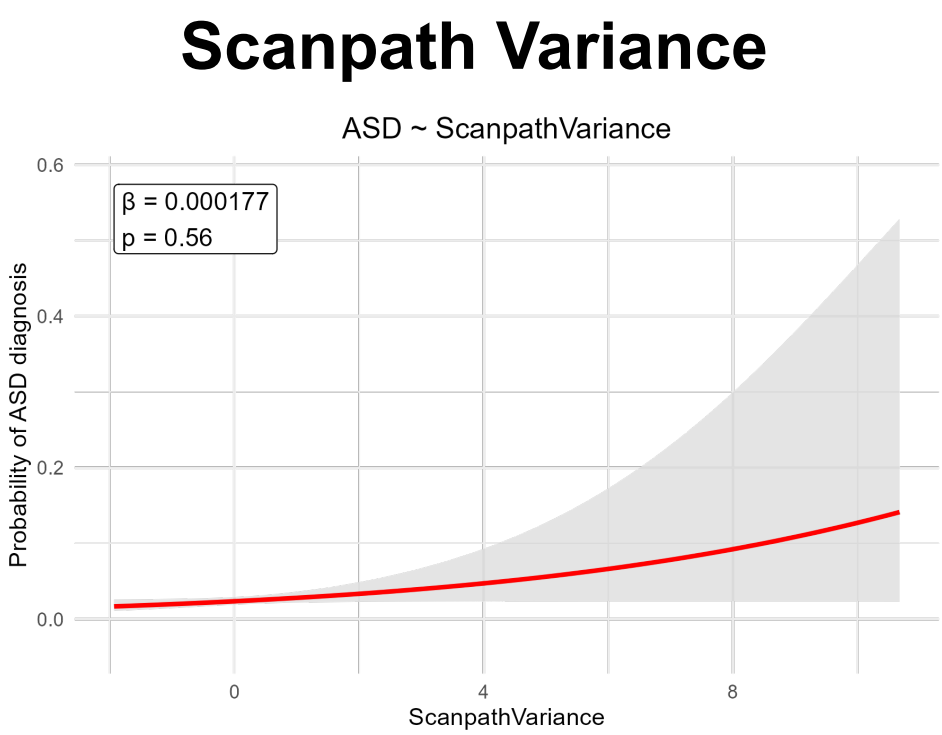


Figure G.11: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Scanpath Variance in male participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

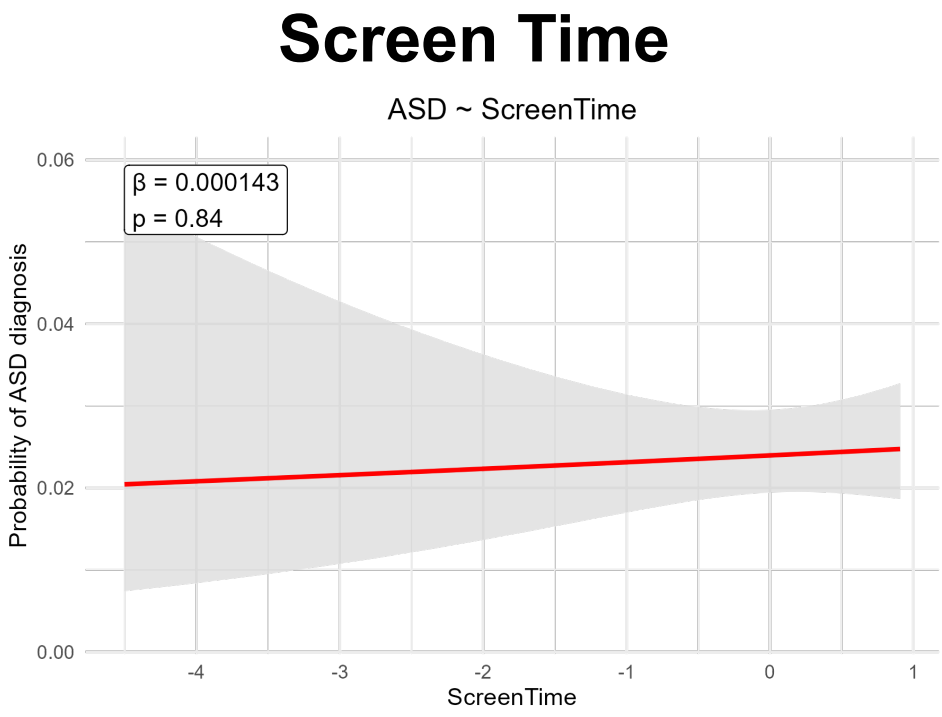


Figure G.12: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Screen Time in male participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

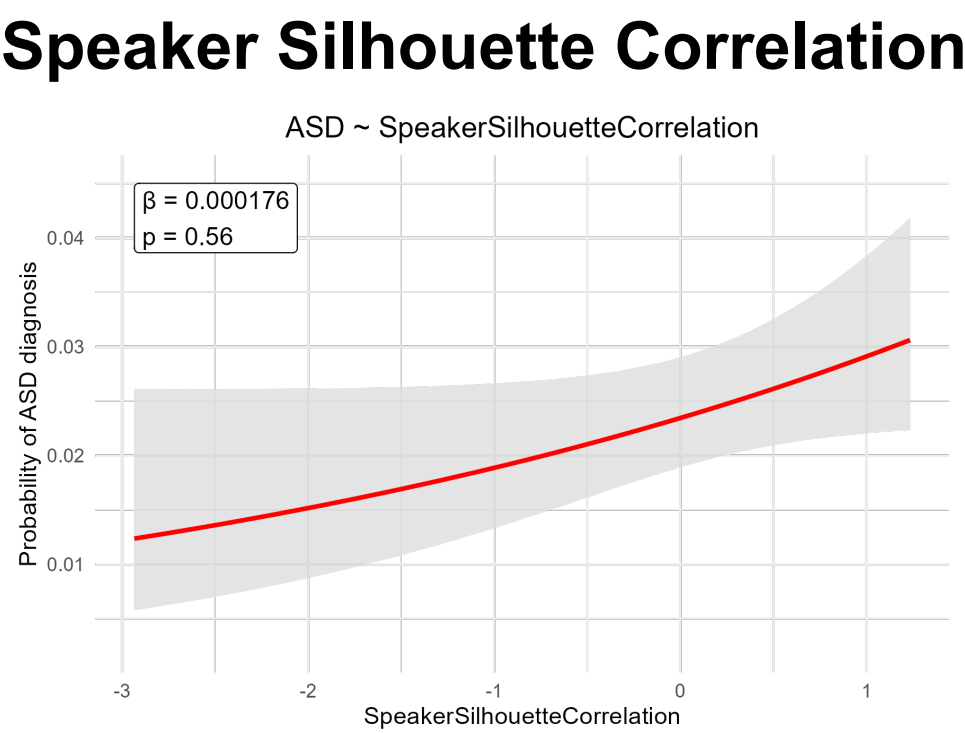


Figure G.13: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Speaker Silhouette Correlation in male participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

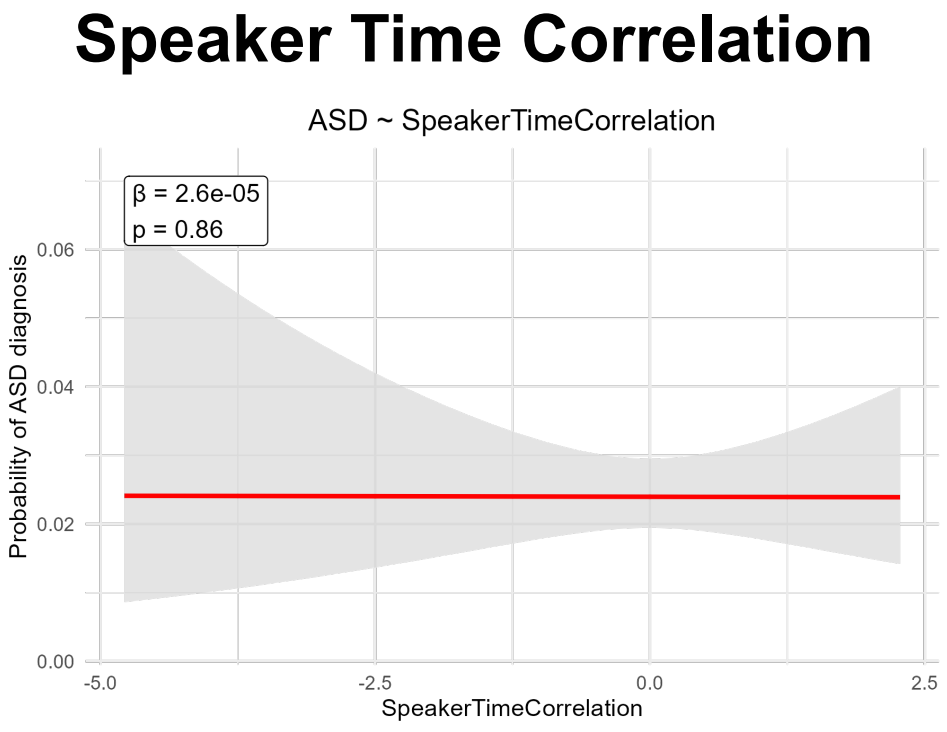


Figure G.14: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Speaker Time Correlation in male participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

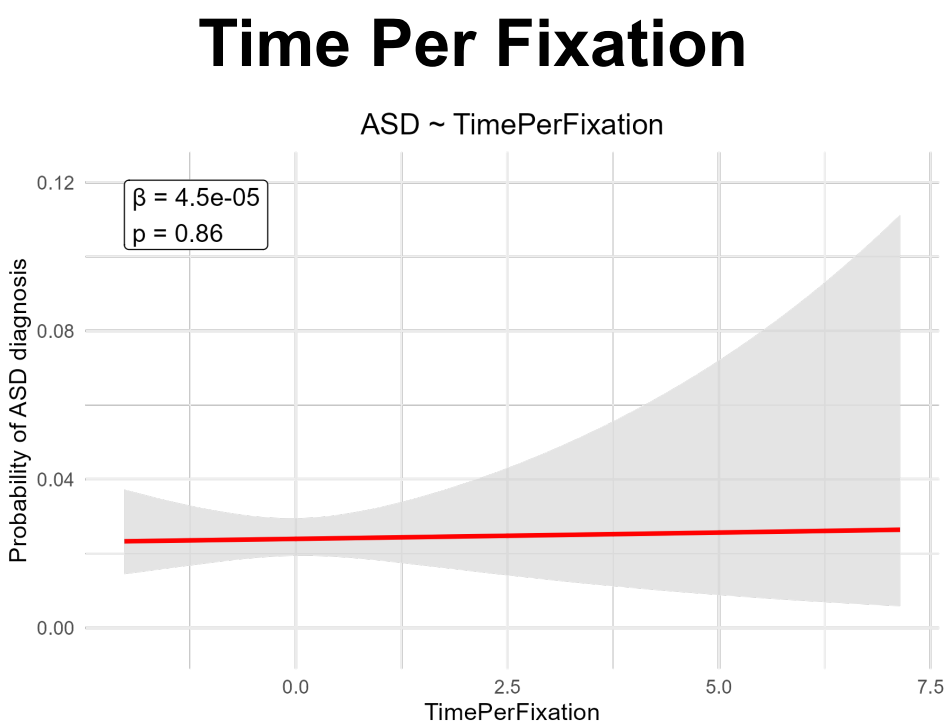


Figure G.15: GEE-GLM model fit illustrating the relationship between ASD diagnosis and Time Per Fixation in male participants. The feature values on the x-axis are presented in z-scores. The ASD diagnosis on the y-axis indicates a probability between zero and one, with the mean being low due to class imbalance. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

H

GEE-GLM SRS Scores: All Participants

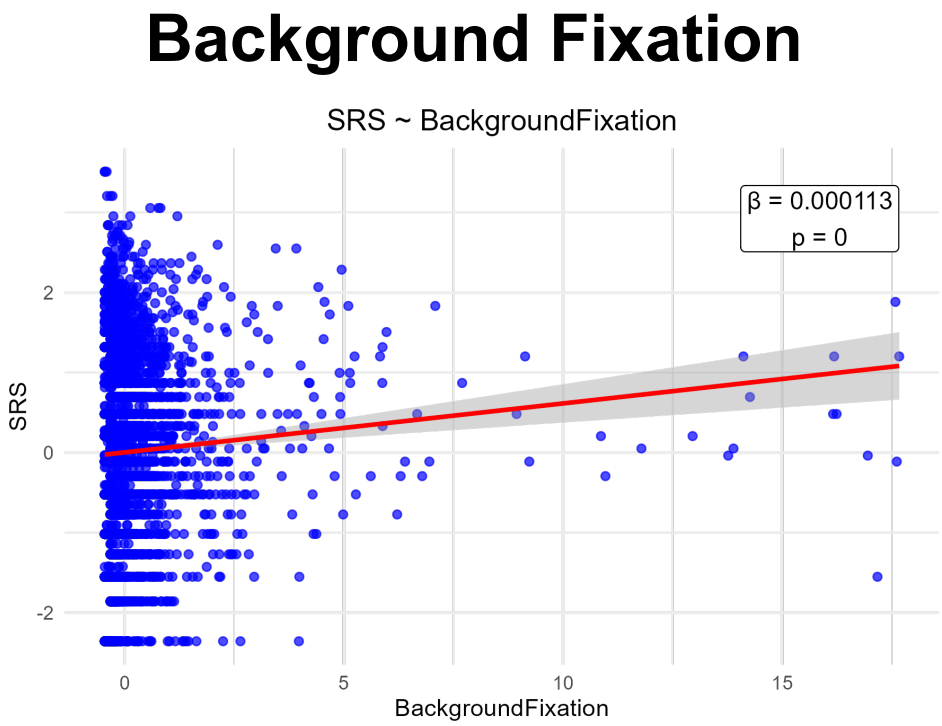


Figure H.1: Scatter plot with GEE-GLM model fit illustrating the relationship between SRS scores and Background Fixation. Each point represents an individual measurement. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores, with mean zero and standard deviation one. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

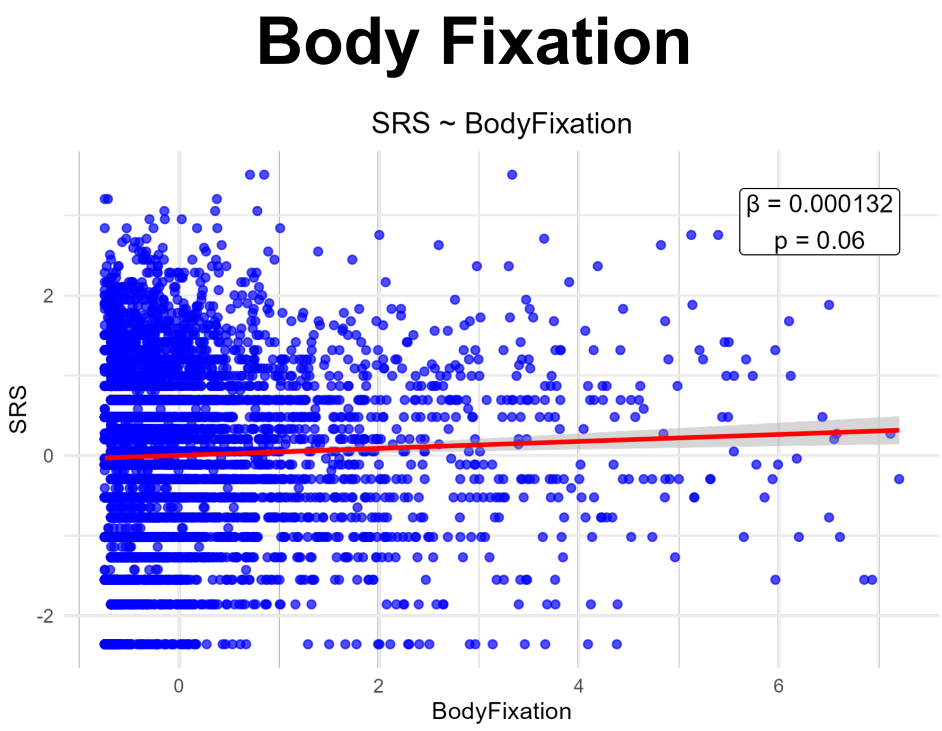


Figure H.2: Scatter plot with GEE-GLM model fit illustrating the relationship between SRS scores and Body Fixation. Each point represents an individual measurement. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores, with mean zero and standard deviation one. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

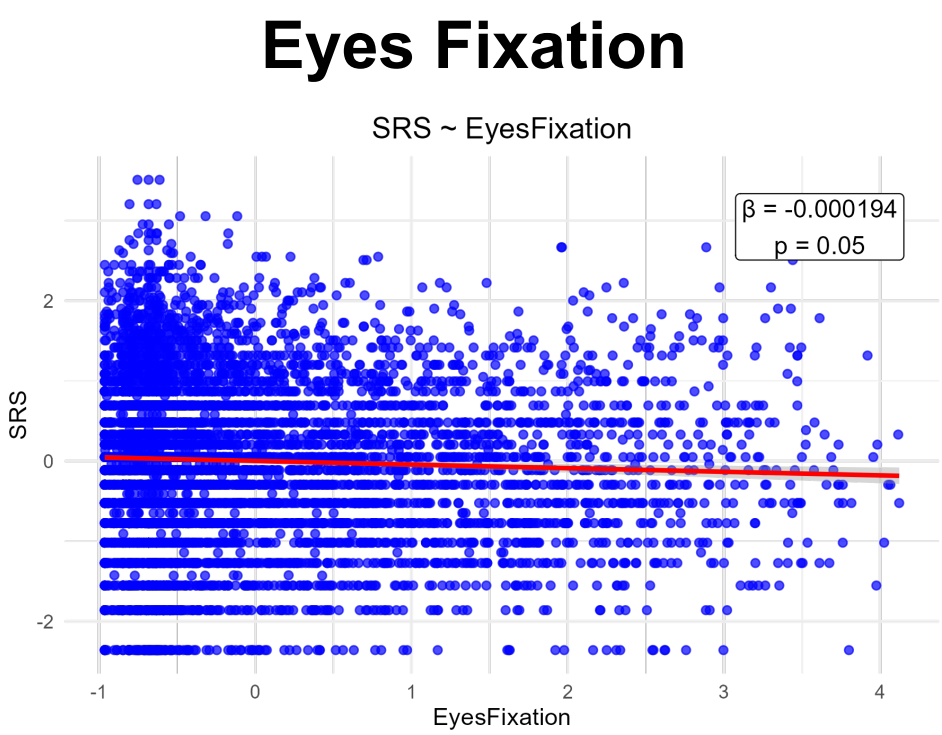


Figure H.3: Scatter plot with GEE-GLM model fit illustrating the relationship between SRS scores and Eyes Fixation. Each point represents an individual measurement. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores, with mean zero and standard deviation one. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

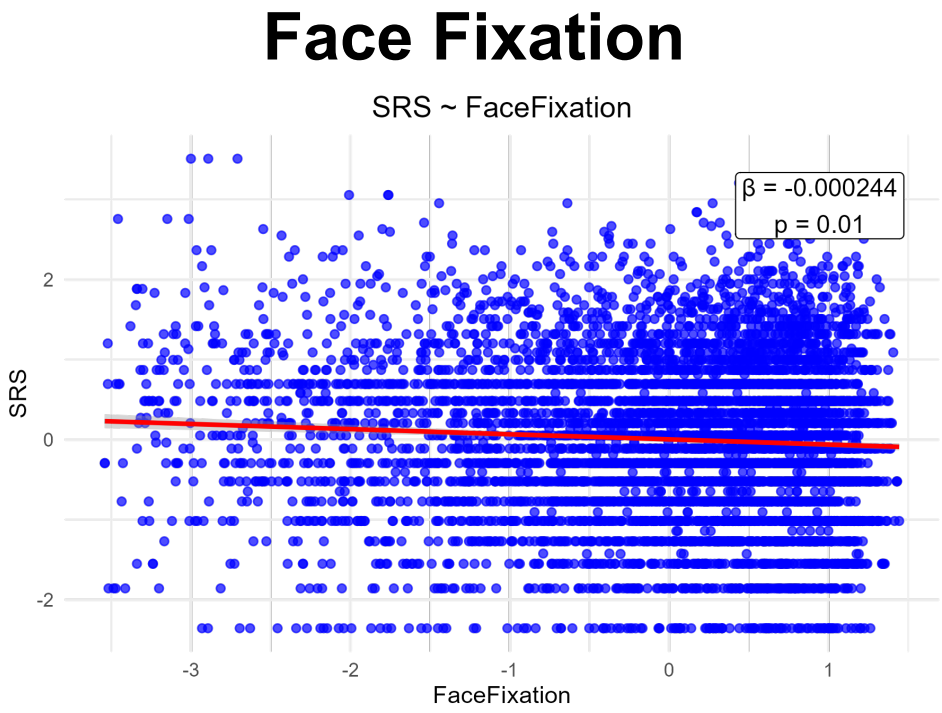


Figure H.4: Scatter plot with GEE-GLM model fit illustrating the relationship between SRS scores and Face Fixation. Each point represents an individual measurement. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores, with mean zero and standard deviation one. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

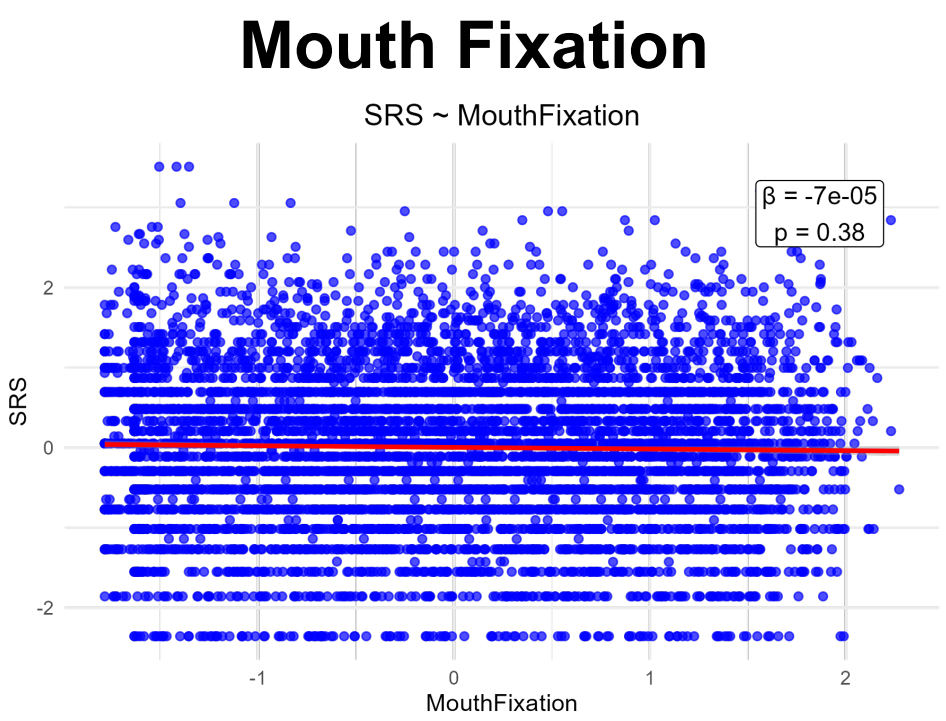


Figure H.5: Scatter plot with GEE-GLM model fit illustrating the relationship between SRS scores and Mouth Fixation. Each point represents an individual measurement. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores, with mean zero and standard deviation one. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

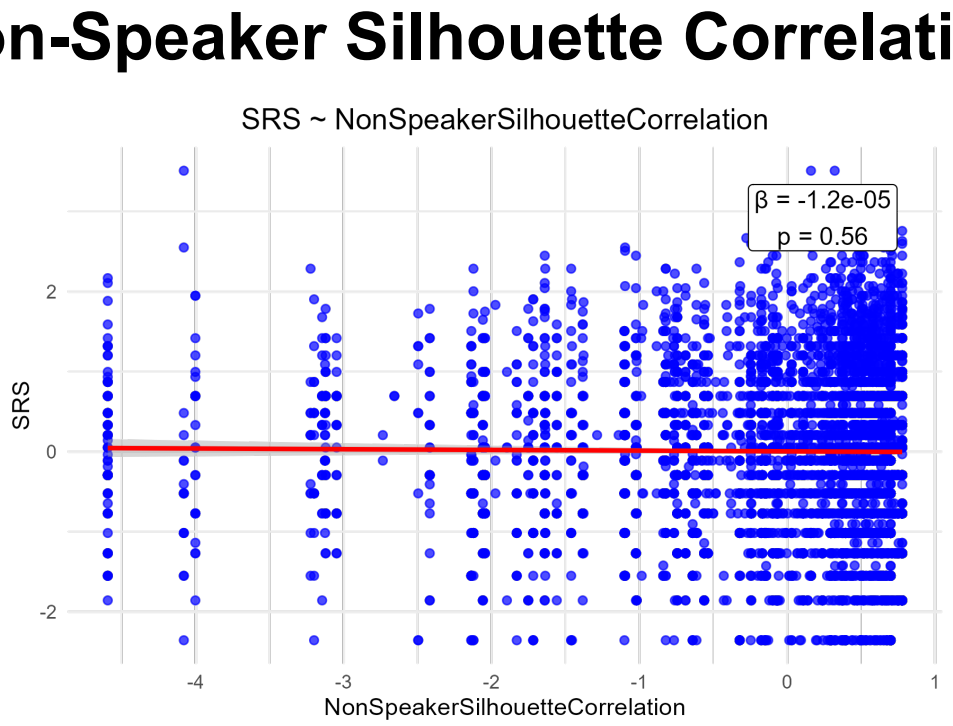


Figure H.6: Scatter plot with GEE-GLM model fit illustrating the relationship between SRS scores and Non-Speaker Silhouette Correlation. Each point represents an individual measurement. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores, with mean zero and standard deviation one. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

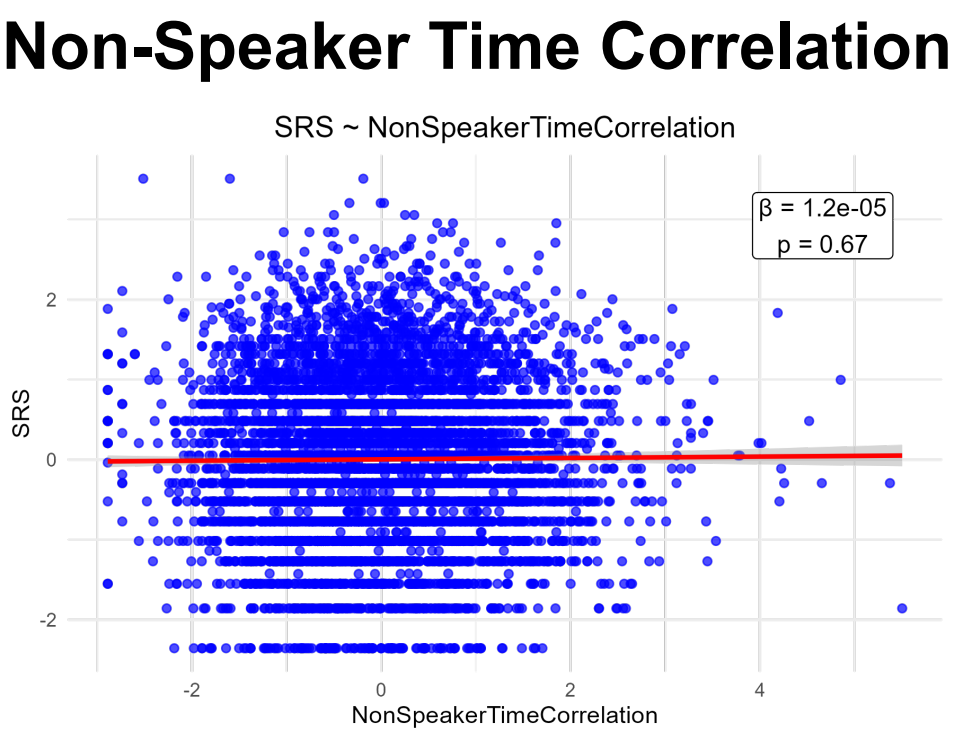


Figure H.7: Scatter plot with GEE-GLM model fit illustrating the relationship between SRS scores and Non-Speaker Time Correlation. Each point represents an individual measurement. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores, with mean zero and standard deviation one. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

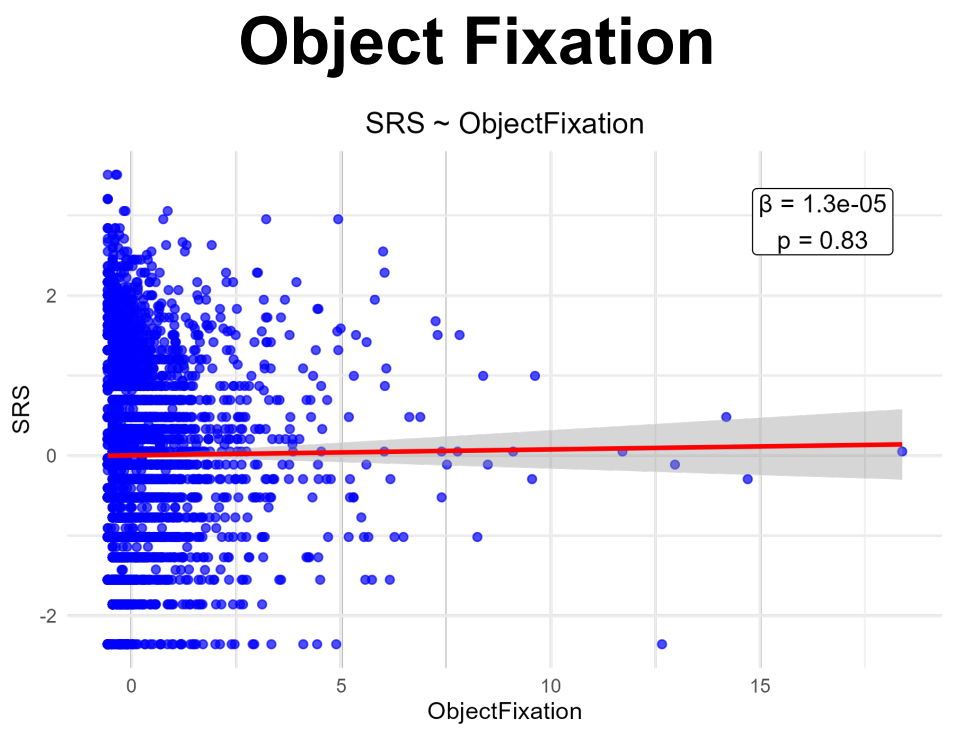


Figure H.8: Scatter plot with GEE-GLM model fit illustrating the relationship between SRS scores and Object Fixation. Each point represents an individual measurement. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores, with mean zero and standard deviation one. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

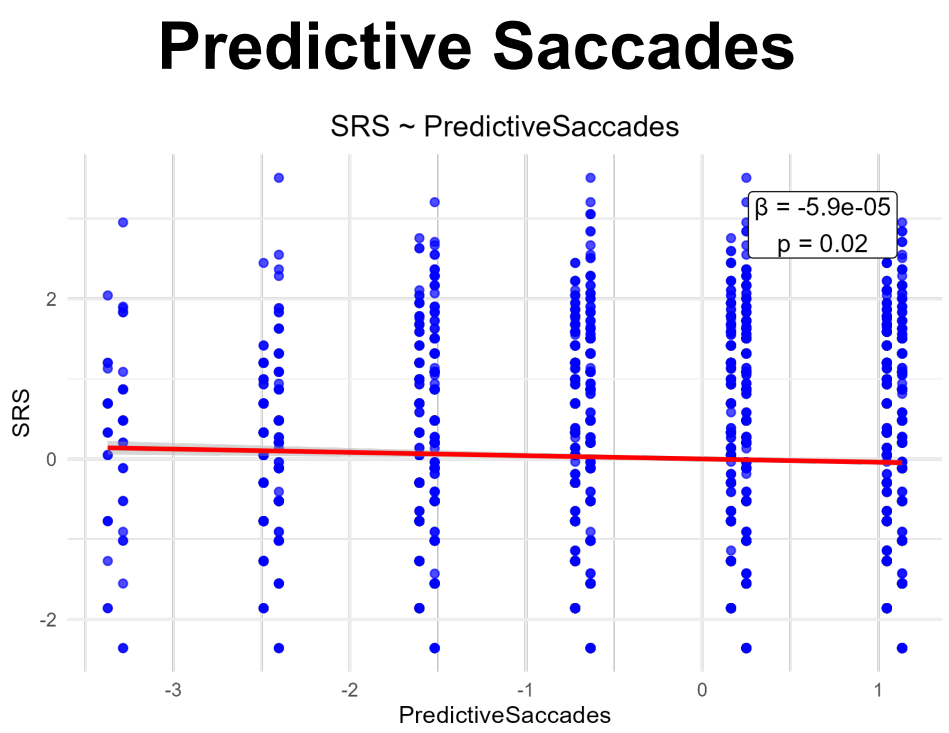


Figure H.9: Scatter plot with GEE-GLM model fit illustrating the relationship between SRS scores and Predictive Saccades. Each point represents an individual measurement. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores, with mean zero and standard deviation one. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

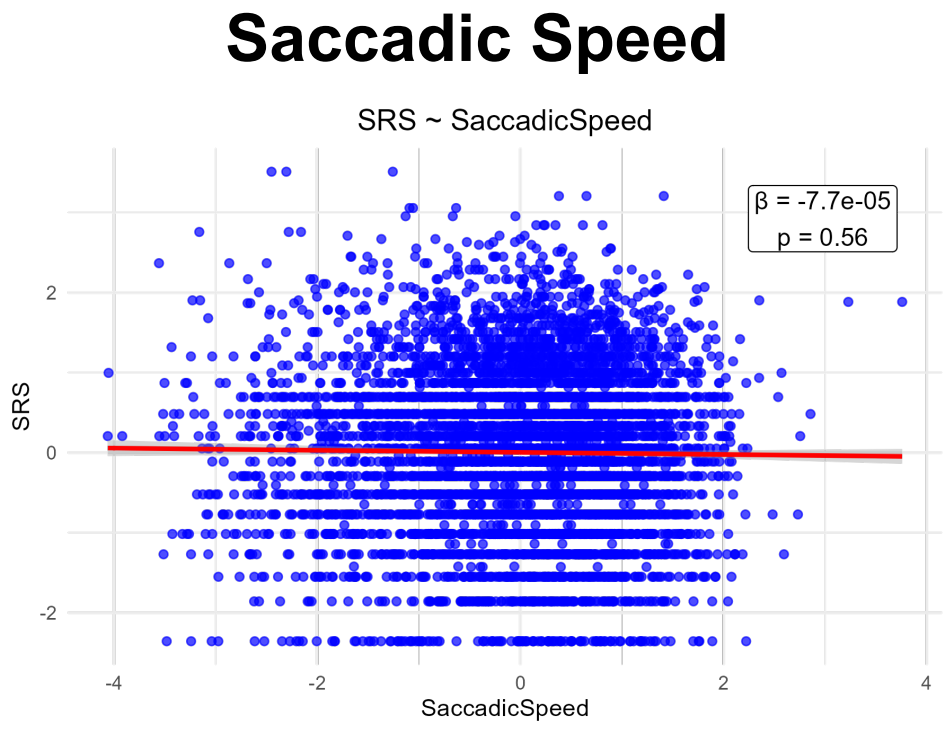


Figure H.10: Scatter plot with GEE-GLM model fit illustrating the relationship between SRS scores and Saccadic Speed. Each point represents an individual measurement. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores, with mean zero and standard deviation one. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

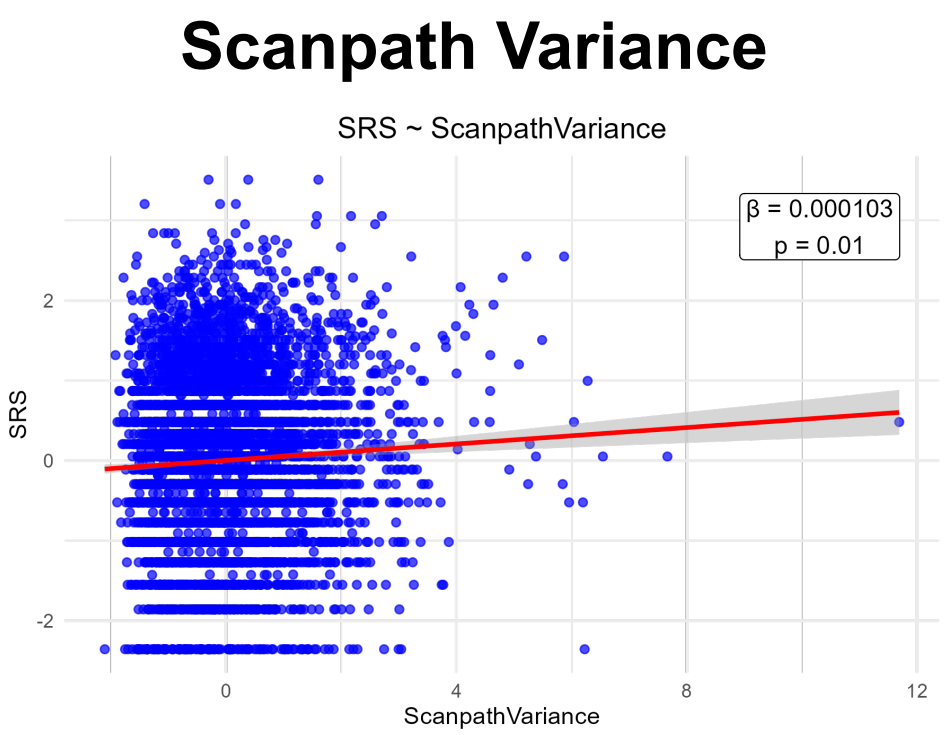


Figure H.11: Scatter plot with GEE-GLM model fit illustrating the relationship between SRS scores and Scanpath Variance. Each point represents an individual measurement. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores, with mean zero and standard deviation one. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

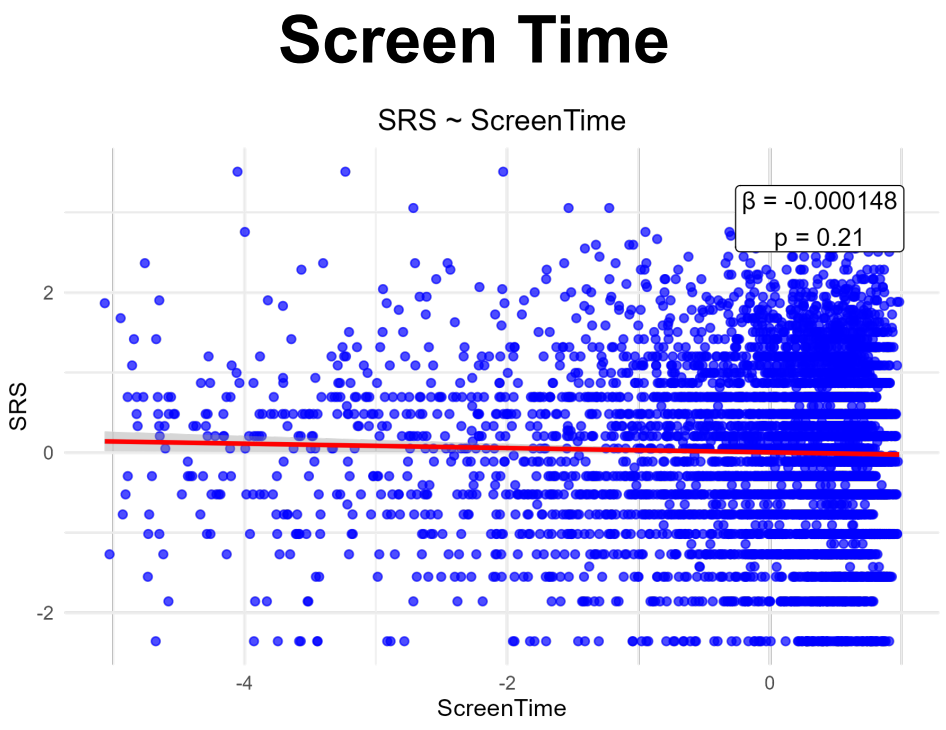


Figure H.12: Scatter plot with GEE-GLM model fit illustrating the relationship between SRS scores and Screen Time. Each point represents an individual measurement. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores, with mean zero and standard deviation one. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

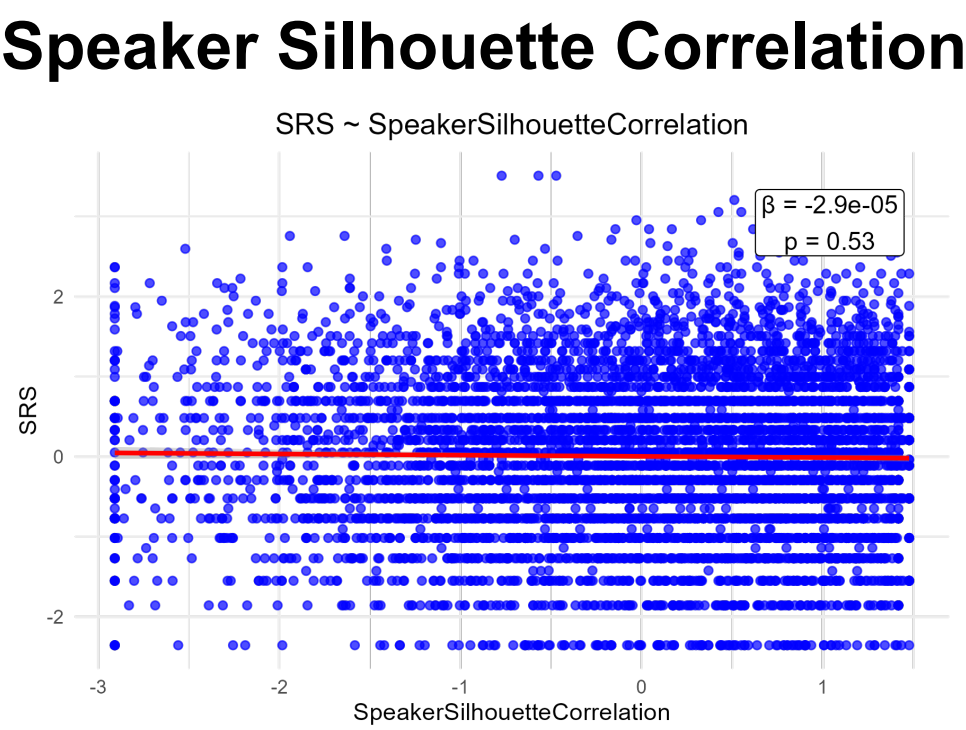


Figure H.13: Scatter plot with GEE-GLM model fit illustrating the relationship between SRS scores and Speaker Silhouette Correlation. Each point represents an individual measurement. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores, with mean zero and standard deviation one. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

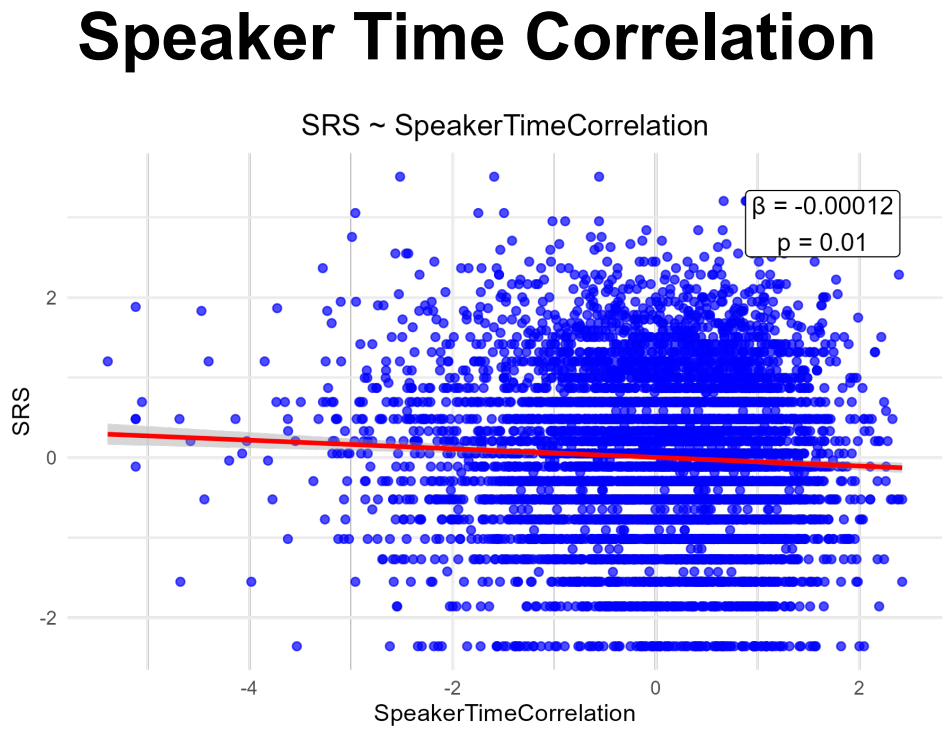


Figure H.14: Scatter plot with GEE-GLM model fit illustrating the relationship between SRS scores and Speaker Time Correlation. Each point represents an individual measurement. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores, with mean zero and standard deviation one. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

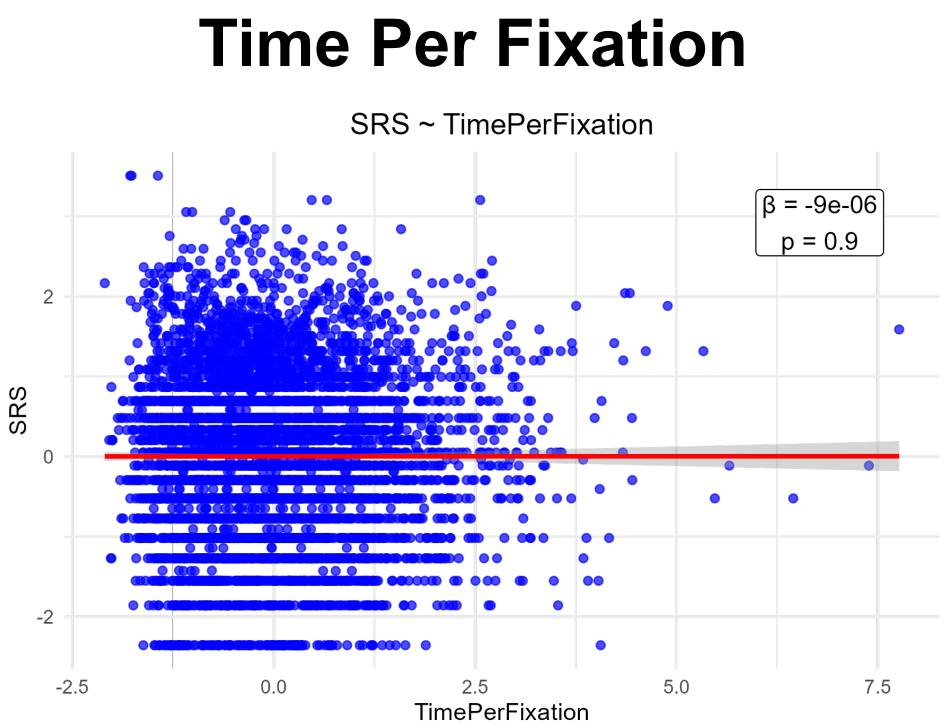


Figure H.15: Scatter plot with GEE-GLM model fit illustrating the relationship between SRS scores and Time Per Fixation. Each point represents an individual measurement. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores, with mean zero and standard deviation one. The red line indicates the model fit, with 95% CI. Coefficient and p-value indicate effect strength and statistical significance.

I

GEE-GLM SRS Scores: Gender Interaction

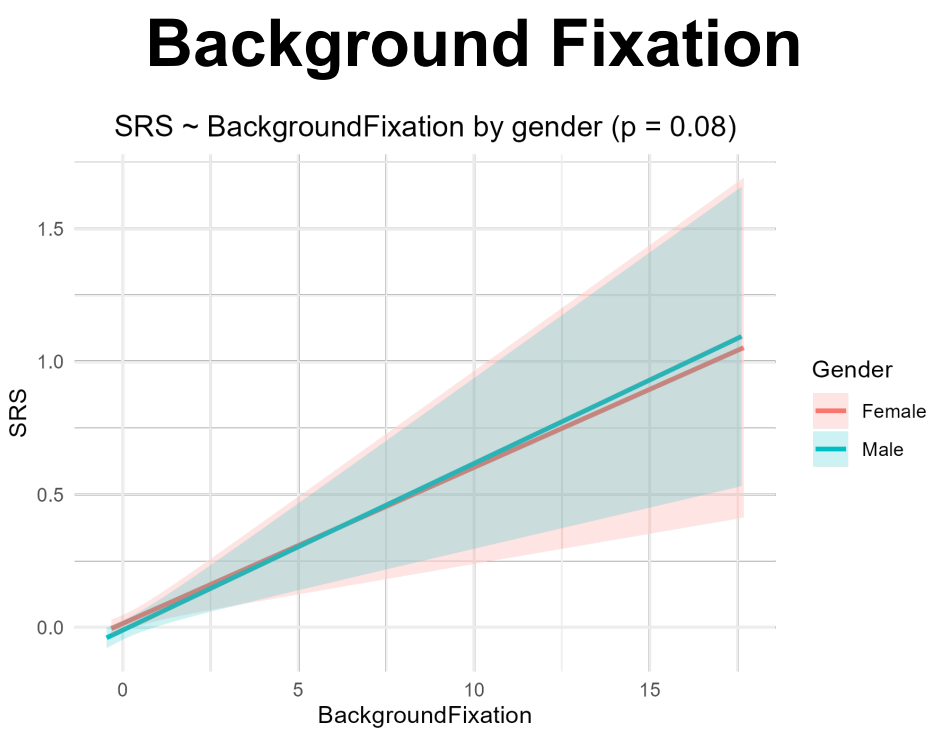


Figure I.1: GEE-GLM model fits illustrating the relationship between SRS scores and Background Fixation for male and female participants. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores. The p-value indicates statistical significance of the difference between male and female participants.

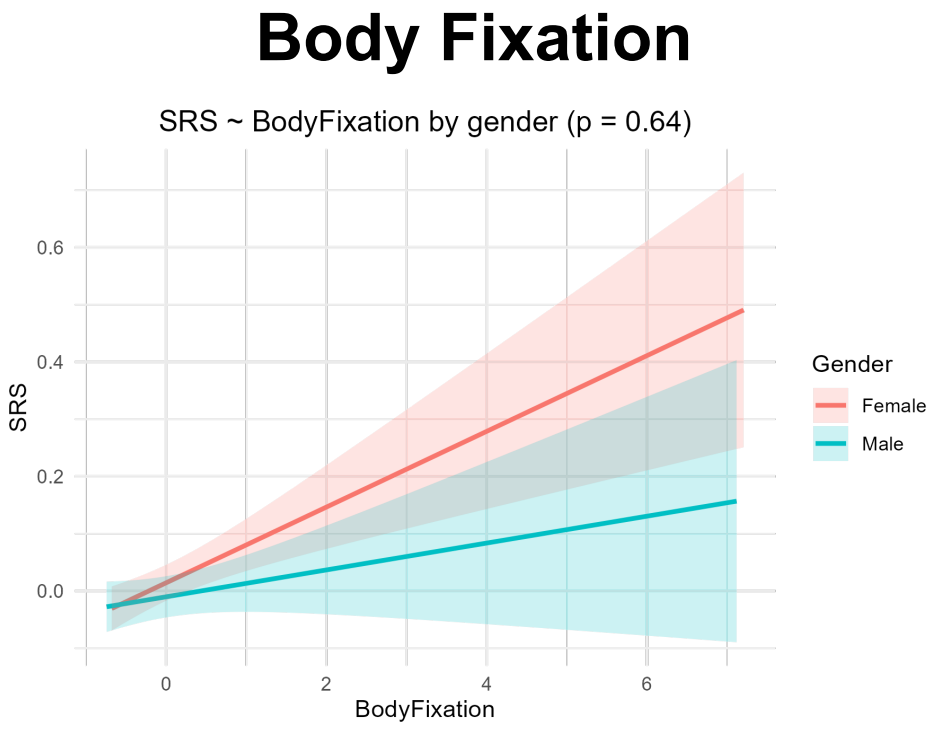


Figure I.2: GEE-GLM model fits illustrating the relationship between SRS scores and Body Fixation for male and female participants. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores. The p-value indicates statistical significance of the difference between male and female participants.

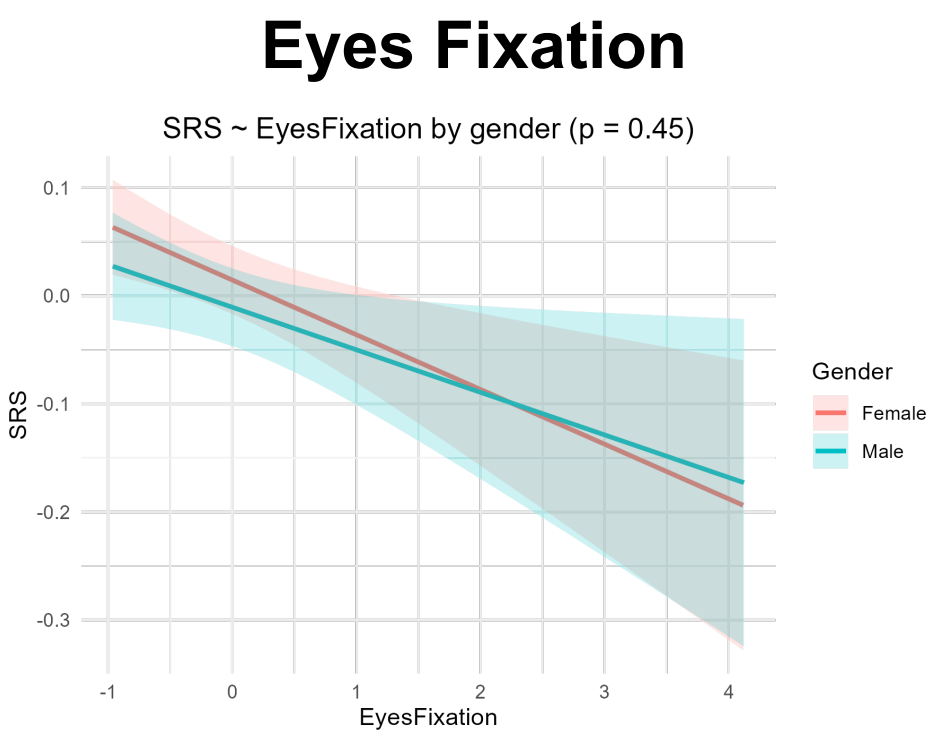


Figure I.3: GEE-GLM model fits illustrating the relationship between SRS scores and Eyes Fixation for male and female participants. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores. The p-value indicates statistical significance of the difference between male and female participants.

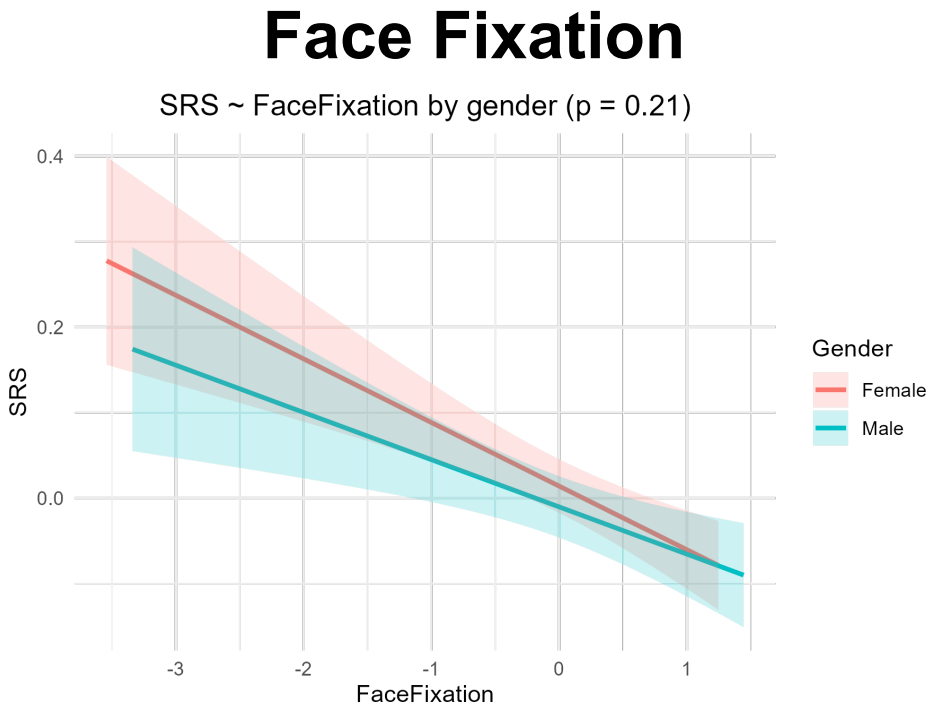


Figure I.4: GEE-GLM model fits illustrating the relationship between SRS scores and Face Fixation for male and female participants. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores. The p-value indicates statistical significance of the difference between male and female participants.

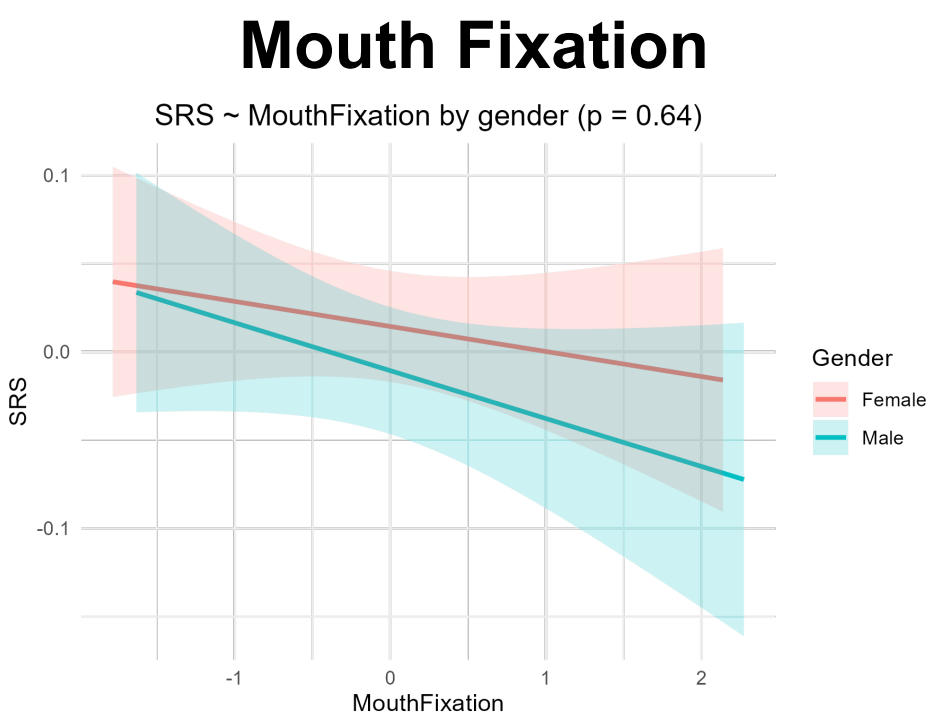


Figure I.5: GEE-GLM model fits illustrating the relationship between SRS scores and Mouth Fixation for male and female participants. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores. The p-value indicates statistical significance of the difference between male and female participants.

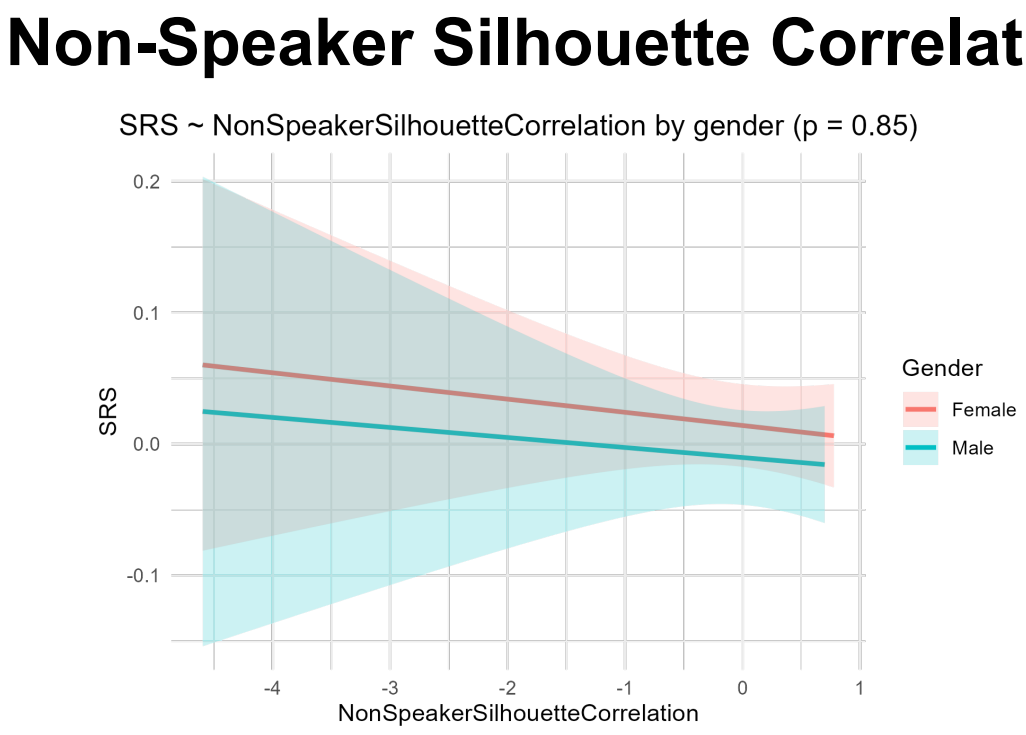


Figure I.6: GEE-GLM model fits illustrating the relationship between SRS scores and Non-Speaker Silhouette Correlation for male and female participants. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores. The p-value indicates statistical significance of the difference between male and female participants.

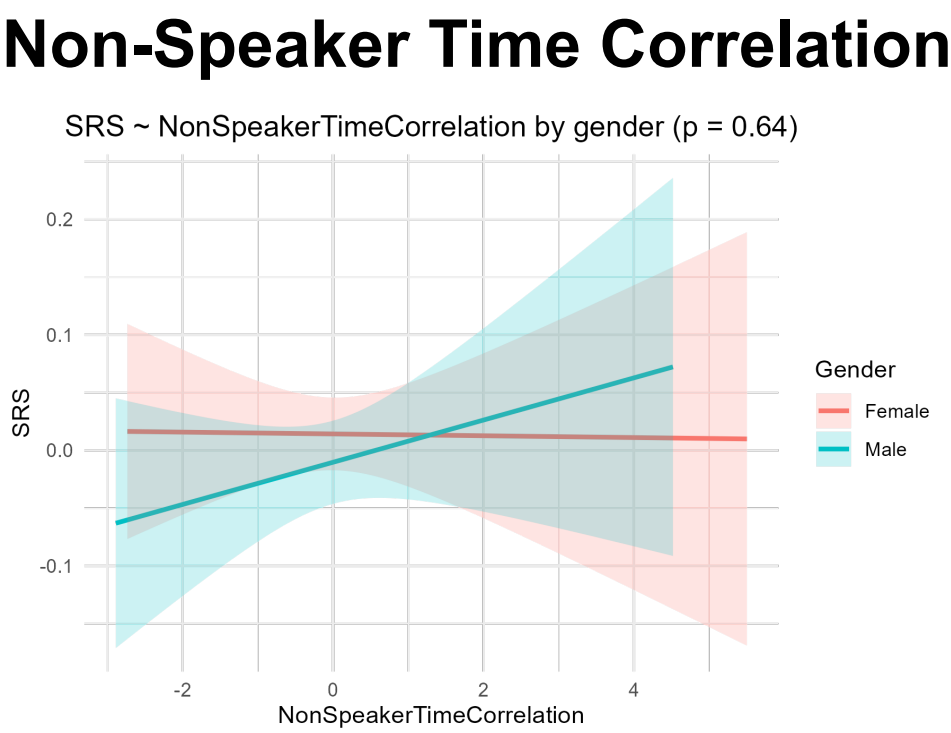


Figure I.7: GEE-GLM model fits illustrating the relationship between SRS scores and Non-Speaker Time Correlation for male and female participants. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores. The p-value indicates statistical significance of the difference between male and female participants.

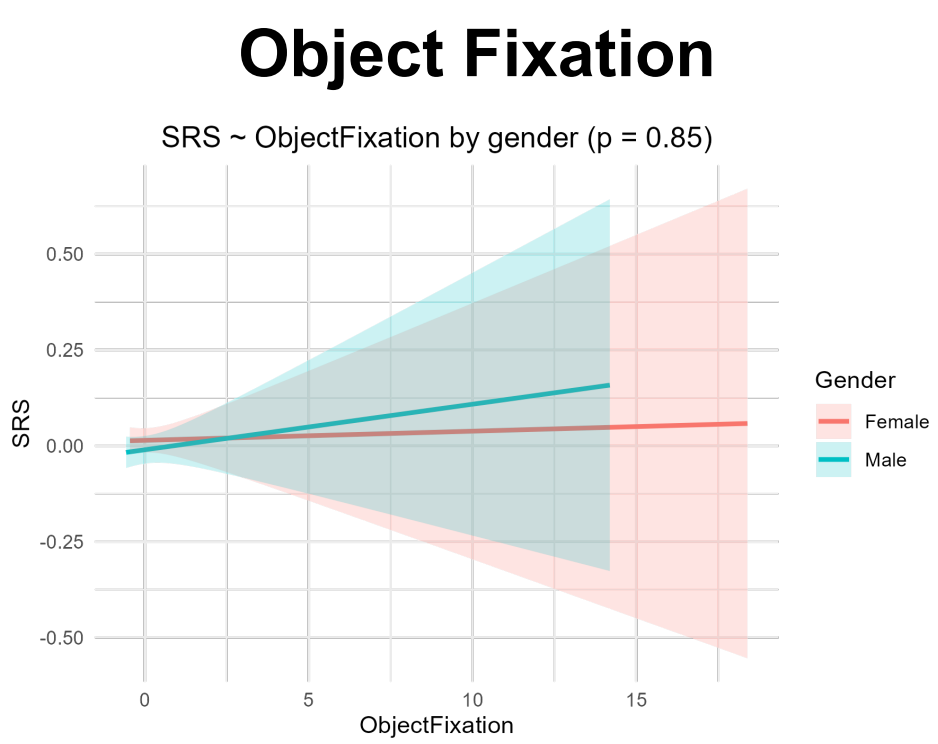


Figure I.8: GEE-GLM model fits illustrating the relationship between SRS scores and Object Fixation for male and female participants. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores. The p-value indicates statistical significance of the difference between male and female participants.

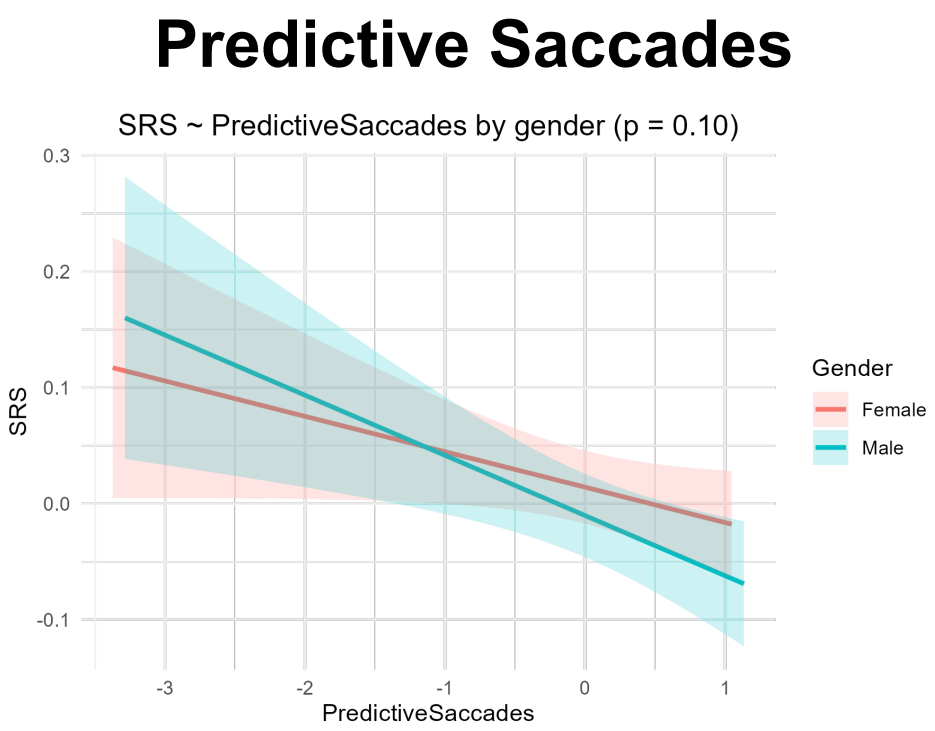


Figure I.9: GEE-GLM model fits illustrating the relationship between SRS scores and Predictive Saccades for male and female participants. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores. The p-value indicates statistical significance of the difference between male and female participants.

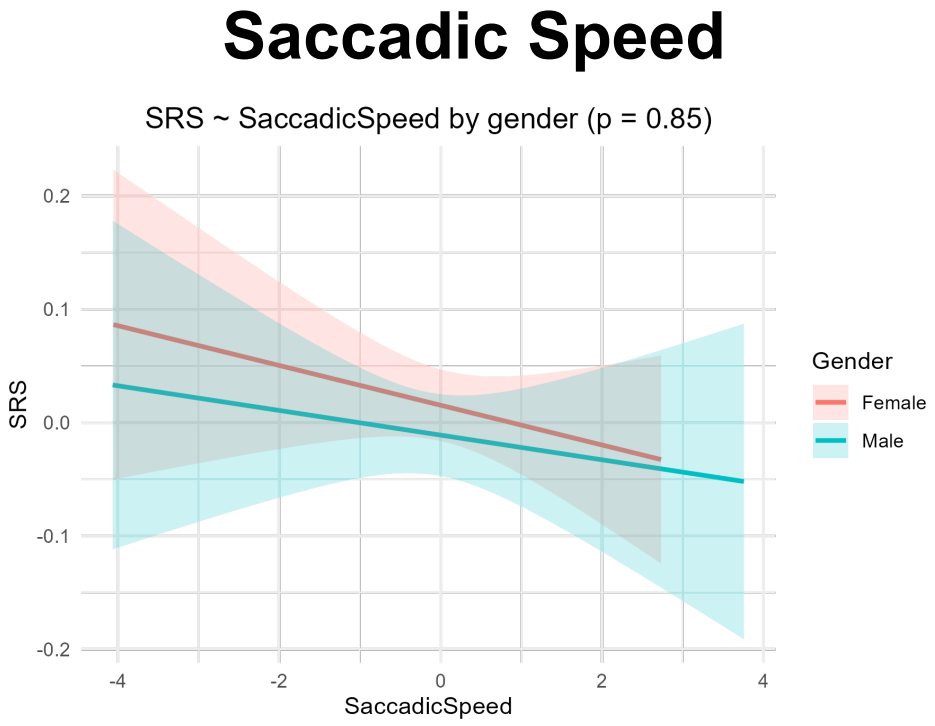


Figure I.10: GEE-GLM model fits illustrating the relationship between SRS scores and Saccadic Speed for male and female participants. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores. The p-value indicates statistical significance of the difference between male and female participants.

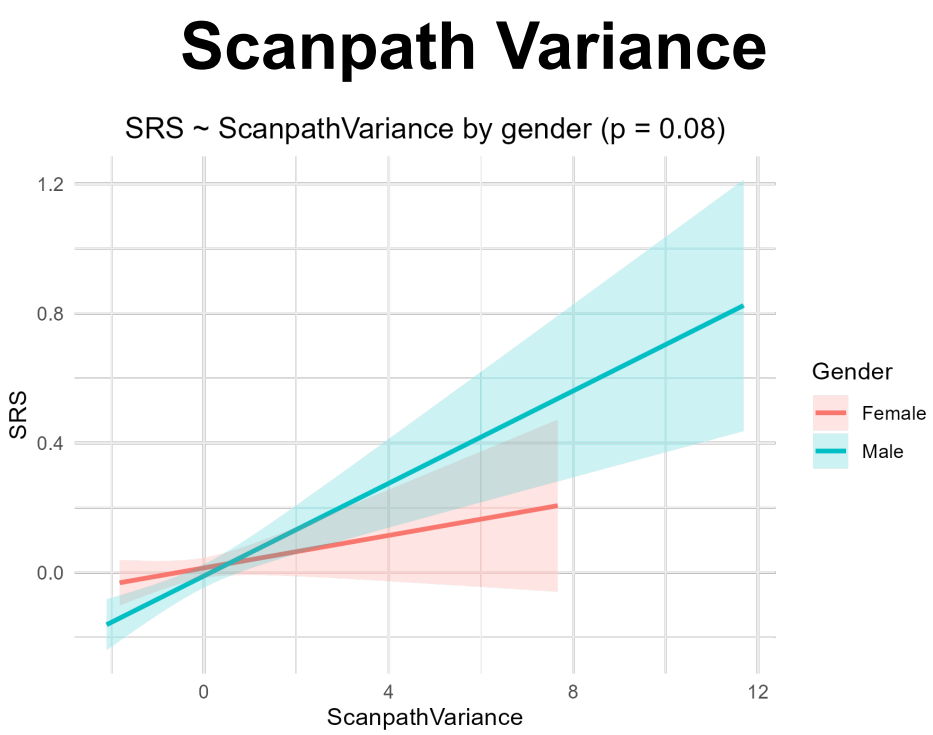


Figure I.11: GEE-GLM model fits illustrating the relationship between SRS scores and Scanpath Variance for male and female participants. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores. The p-value indicates statistical significance of the difference between male and female participants.

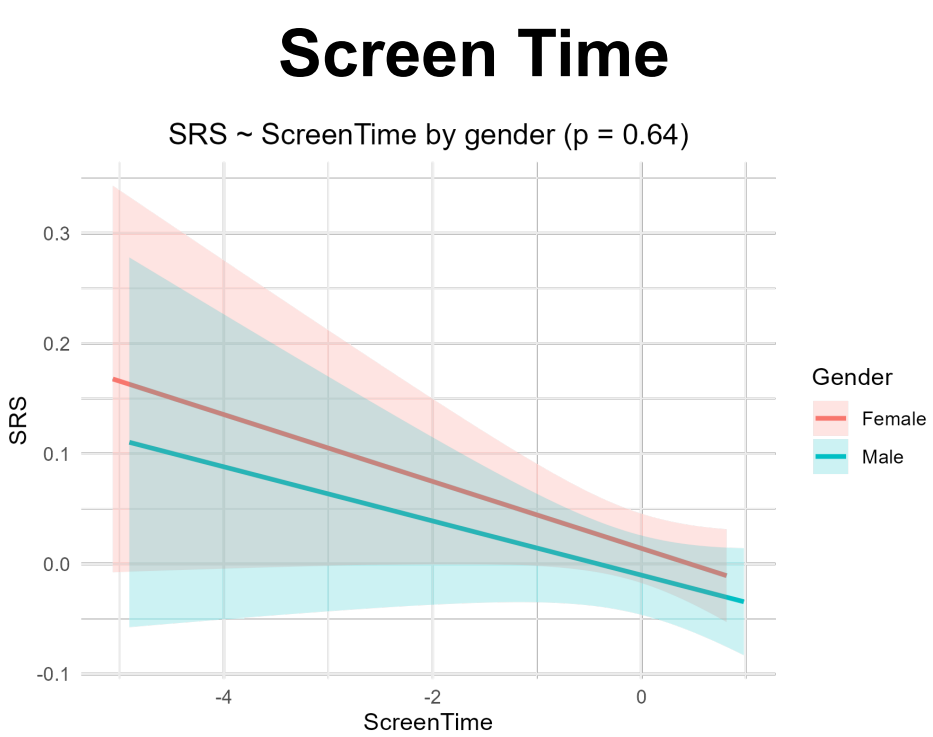


Figure I.12: GEE-GLM model fits illustrating the relationship between SRS scores and Screen Time for male and female participants. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores. The p-value indicates statistical significance of the difference between male and female participants.

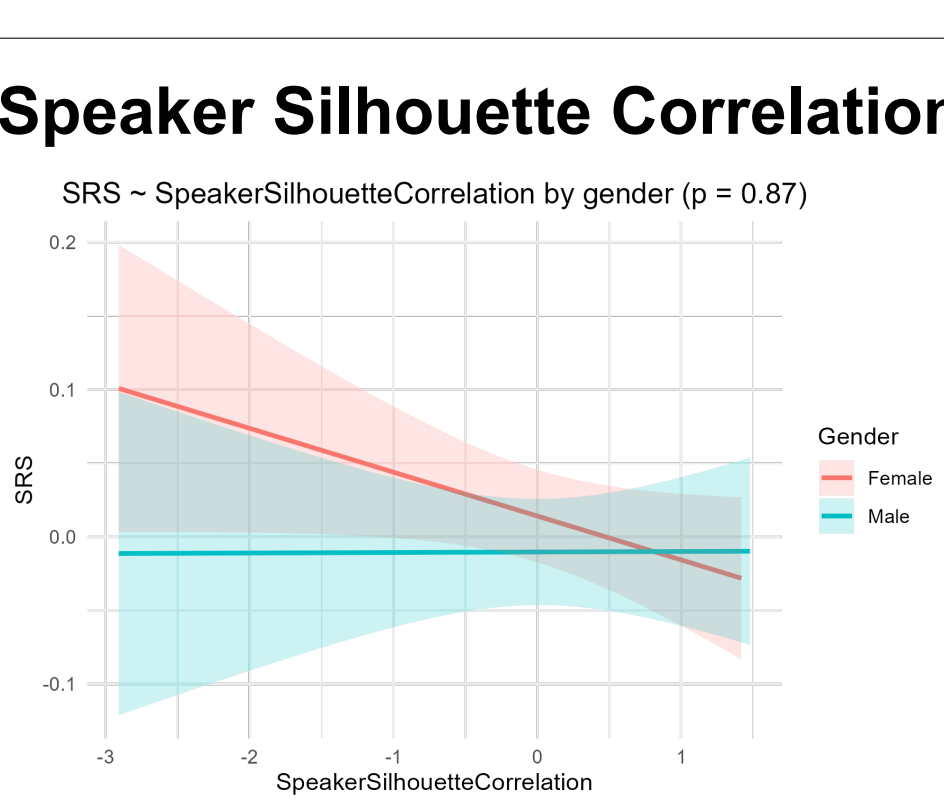


Figure I.13: GEE-GLM model fits illustrating the relationship between SRS scores and Speaker Silhouette Correlation for male and female participants. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores. The p-value indicates statistical significance of the difference between male and female participants.

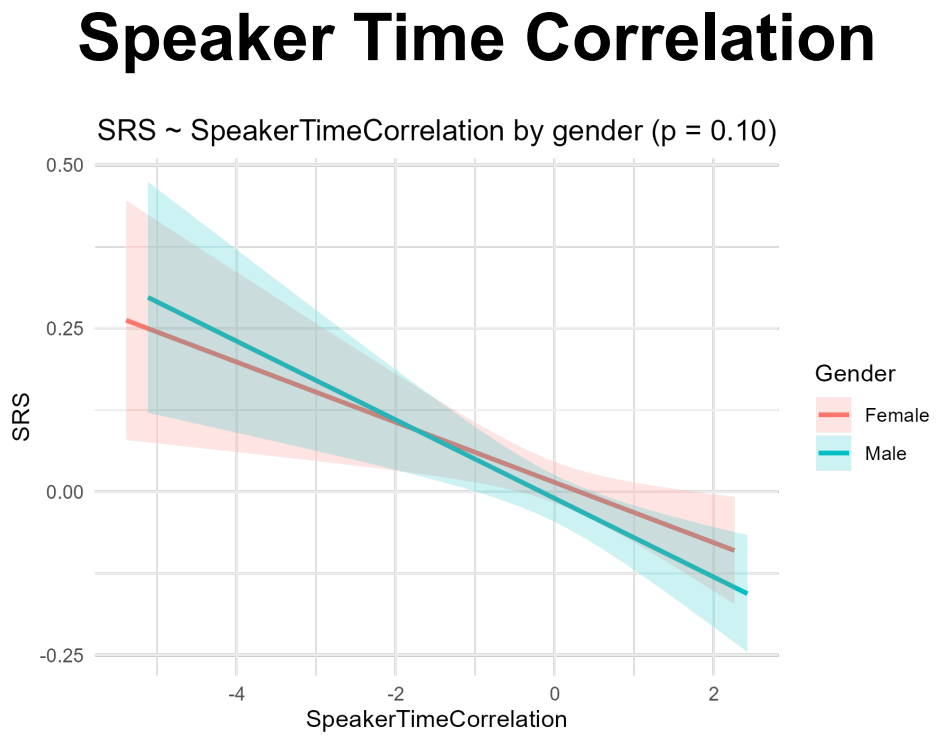


Figure I.14: GEE-GLM model fits illustrating the relationship between SRS scores and Speaker Time Correlation for male and female participants. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores. The p-value indicates statistical significance of the difference between male and female participants.

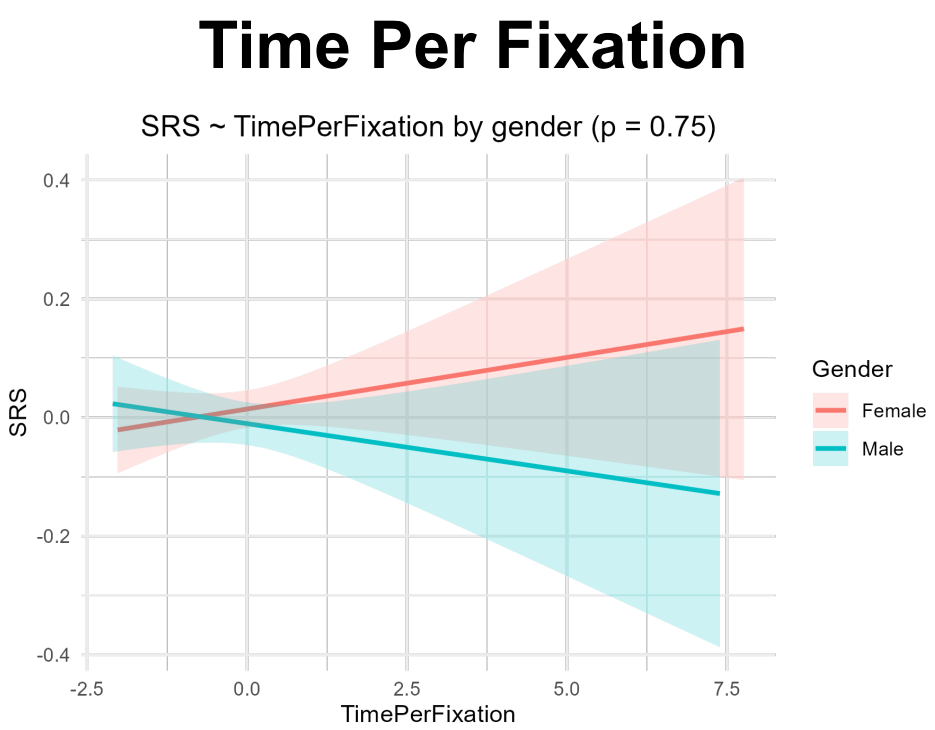


Figure I.15: GEE-GLM model fits illustrating the relationship between SRS scores and Time Per Fixation for male and female participants. The feature values on the x-axis and SRS scores on the y-axis are presented in z-scores. The p-value indicates statistical significance of the difference between male and female participants.

J

Feature Correlation and VIF

ASD Diagnosis

Table J.1: Correlation matrix of the feature set predicting ASD diagnosis. **Abbreviations:** BodyFx = Body Fixation; FaceFx = Face Fixation; NSpTC = Non-Speaker Time Correlation; ObjFx = Object Fixation; PredSac = Predictive Saccades; ScanVar = Scanpath Variance; SpSC = Speaker Silhouette Correlation.

	BodyFx	**FaceFx**	**NSpTC**	**ObjFx**	**PredSac**	**ScanVar**	**SpSC**
BodyFx	1	0.60	0.04	0.00	0.02	0.12	0.01
FaceFx	0.60	1	0.07	0.19	0.29	0.40	0.23
NSpTC	0.04	0.07	1	0.11	0.08	0.41	0.06
ObjFx	0.00	0.19	0.11	1	0.12	0.58	0.04
PredSac	0.02	0.29	0.08	0.12	1	0.17	0.11
ScanVar	0.12	0.40	0.41	0.58	0.17	1	0.04
SpSC	0.01	0.23	0.06	0.04	0.11	0.04	1

Table J.2: VIF of features in the feature set predicting ASD diagnosis, ordered by decreasing VIF.

Feature	**VIF**
Face Fixation	2.87
Scanpath Variance	2.41
Body Fixation	1.90
Object Fixation	1.57
Non-Speaker Time Correlation	1.49
Predictive Saccades	1.25
Speaker Silhouette Correlation	1.22

SRS Scores

Table J.3: Correlation matrix of the feature set predicting SRS scores. **Abbreviations:** BackFx = Background Fixation; FaceFx = Face Fixation; PredSac = Predictive Saccades; ScanVar = Scanpath Variance; SpTC = Speaker Time Correlation.

	BackFx	**FaceFx**	**PredSac**	**ScanVar**	**SpTC**
BackFx	1	0.31	0.23	0.27	0.41
FaceFx	0.31	1	0.29	0.40	0.58
PredSac	0.23	0.29	1	0.17	0.36
ScanVar	0.27	0.40	0.17	1	0.50
SpTC	0.41	0.58	0.36	0.50	1

Table J.4: VIF of features in the feature set predicting SRS scores, ordered by decreasing VIF.

Feature	**VIF**
Speaker Time Correlation	1.93
Face Fixation	1.58
Scanpath Variance	1.37
Background Fixation	1.23
Predictive Saccades	1.17

K

Hyperparameter Tuning

ASD Diagnosis

Table K.1: Selected hyperparameters across outer folds for each classifier tried for ASD predictive modelling. The most frequently selected value across folds is shown in the last column.

Model	Hyperparameter	Fold 1	Fold 2	Fold 3	most chosen
Logistic Regression	C	0.464	0.022	0.022	0.022
	penalty	elasticnet	l2	elasticnet	elasticnet
	l1_ratio	0.2	—	0.2	0.2
Random Forest	max_depth	20	5	20	20
	max_features	sqrt	sqrt	sqrt	sqrt
	n_estimators	50	150	50	50
CatBoost	depth	4	4	8	4
	iterations	200	200	300	200
	l2_leaf_reg	1	3	1	1
	learning_rate	0.1	0.05	0.1	0.1
Logistic Regression (PCA)	C	0.022	0.464	0.001	—
	penalty	elasticnet	l1	l2	—
	l1_ratio	0.2	—	—	0.2
	n_components	2	4	2	2
Logistic Regression (LDA)	C	0.001	0.001	0.001	0.001
	penalty	l2	l2	l2	l2
	l1_ratio	—	—	—	—
Ensemble			NA		

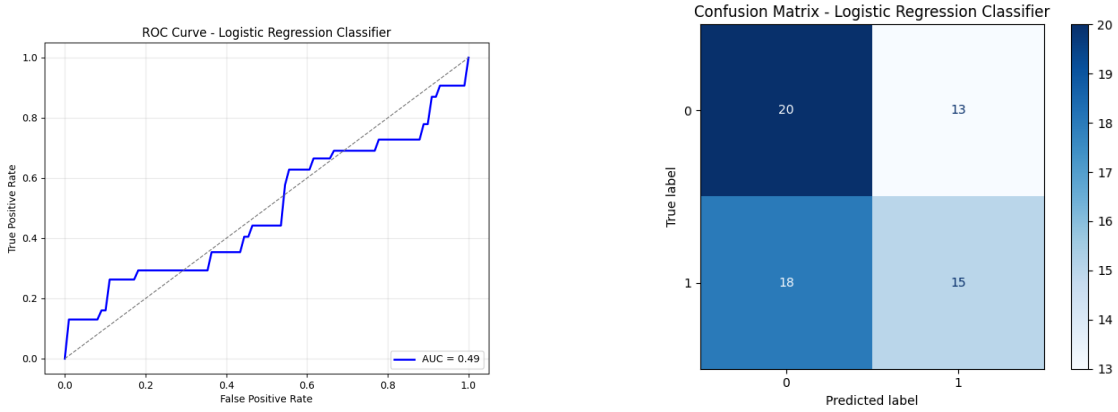
SRS Scores

Table K.2: Selected hyperparameters across outer folds for each classifier tried for SRS predictive modelling. The most frequently selected value across folds is shown in the last column.

Model	Hyperparameter	Fold 1	Fold 2	Fold 3	most chosen
Logistic Regression	C	0.022	0.001	10.0	—
	penalty	elasticnet	l1	elasticnet	elasticnet
	l1_ratio	0.2	—	0.5	—
Random Forest	max_depth	20	5	5	5
	max_features	sqrt	None	sqrt	sqrt
	n_estimators	50	100	50	50
CatBoost	depth	8	8	4	8
	iterations	100	200	100	100
	l2_leaf_reg	5	1	1	1
	learning_rate	0.05	0.01	0.01	0.01
Logistic Regression (PCA)	C	10.0	10.0	10.0	10.0
	penalty	l2	elasticnet	l2	l2
	l1_ratio	—	0.2	—	—
	n_components	4	4	6	4
Logistic Regression (LDA)	C	0.001	0.001	0.001	0.001
	penalty	l2	l1	l2	l2
	l1_ratio	—	—	—	—
Ensemble			NA		

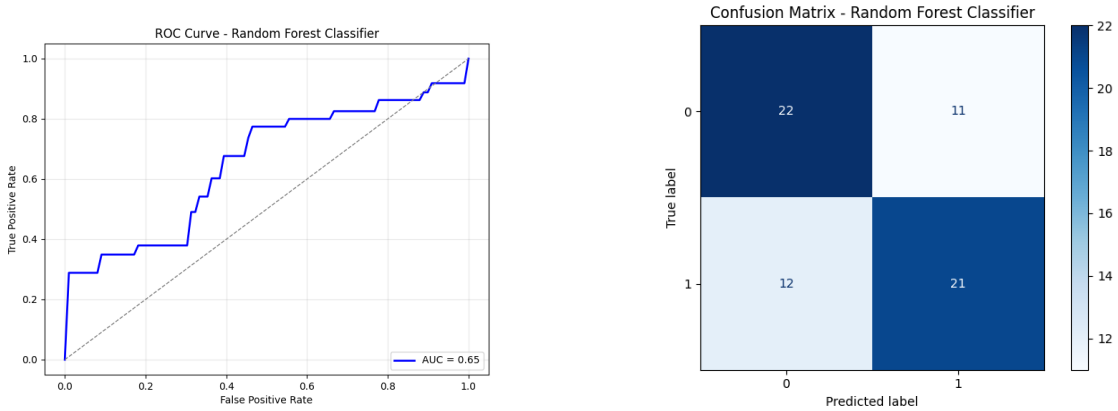
L

ASD Models



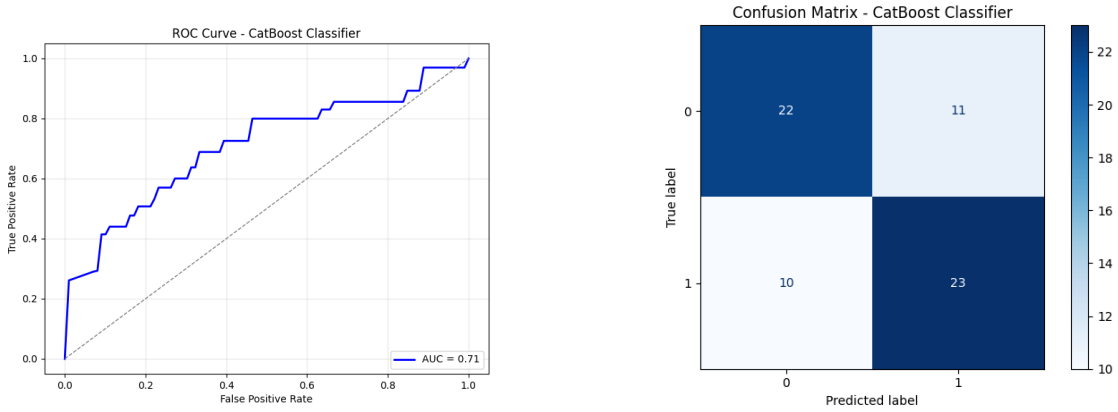
(a) ROC curve **(b)** Confusion matrix

Figure L.1: Evaluation metrics for the Logistic Regression classifier: AUC = 0.49; F1 = 0.49; precision = 0.56; recall = 0.46.



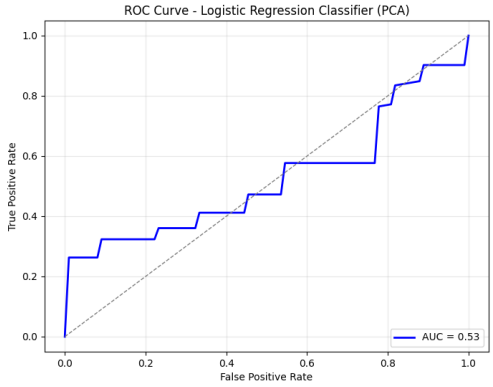
(a) ROC curve **(b)** Confusion matrix

Figure L.2: Evaluation metrics for the Random Forest classifier: AUC = 0.65; F1 = 0.64; precision = 0.69; recall = 0.66.

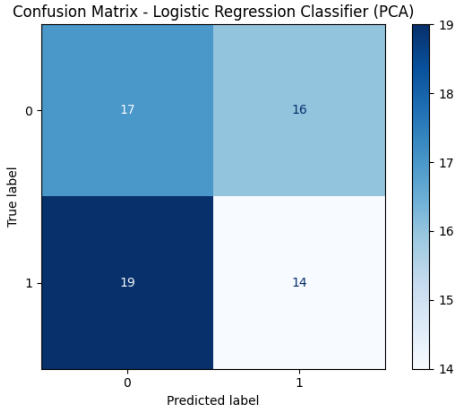


(a) ROC curve **(b)** Confusion matrix

Figure L.3: Evaluation metrics for the CatBoost Gradient Boosting Decision Tree classifier: AUC = 0.71; F1 = 0.69; precision = 0.69; recall = 0.72.

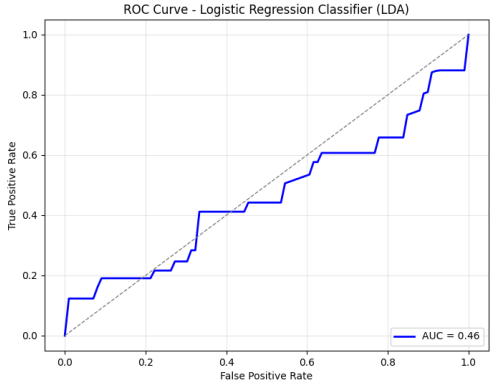


(a) ROC curve

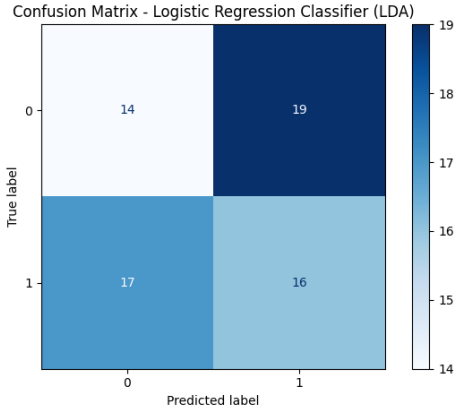


(b) Confusion matrix

Figure L.4: Evaluation metrics for the Logistic Regression classifier with PCA: AUC = 0.53; F1 = 0.45; precision = 0.54; recall = 0.43.

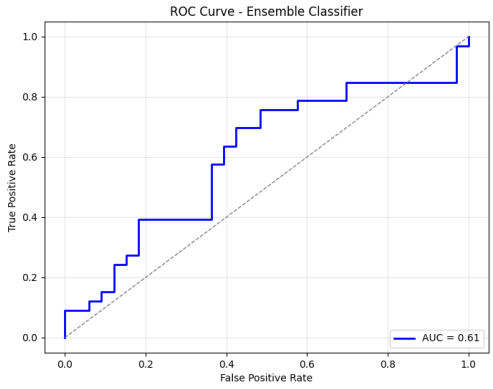


(a) ROC curve

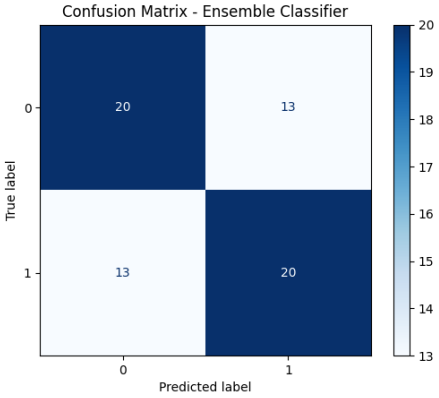


(b) Confusion matrix

Figure L.5: Evaluation metrics for the Logistic Regression classifier with LDA: AUC = 0.46; F1 = 0.46; precision = 0.46; recall = 0.49.



(a) ROC curve

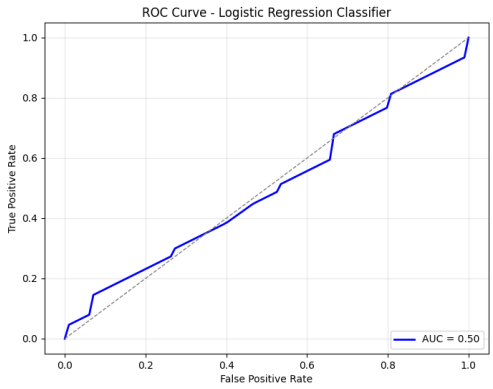


(b) Confusion matrix

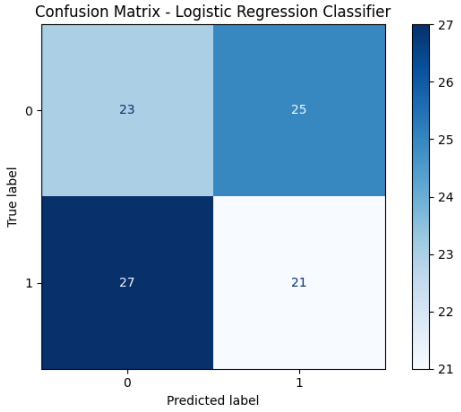
Figure L.6: Evaluation metrics for the Ensemble classifier averaging the Logistic Regression classifier, the Random Forest classifier and the CatBoost Gradient Boosting Decision Tree classifier: AUC = 0.61; F1 = 0.61; precision = 0.61; recall = 0.61.

M

SRS Models

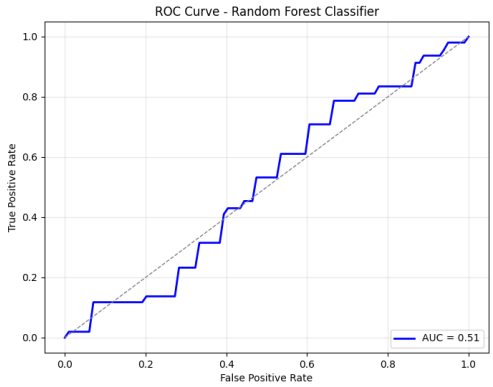


(a) ROC curve

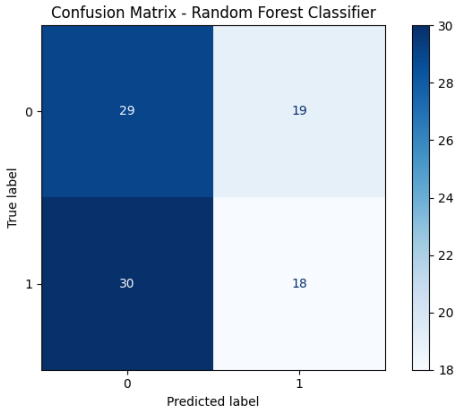


(b) Confusion matrix

Figure M.1: Evaluation metrics for the Logistic Regression classifier: AUC = 0.50; F1 = 0.35; precision = 0.31; recall = 0.47.

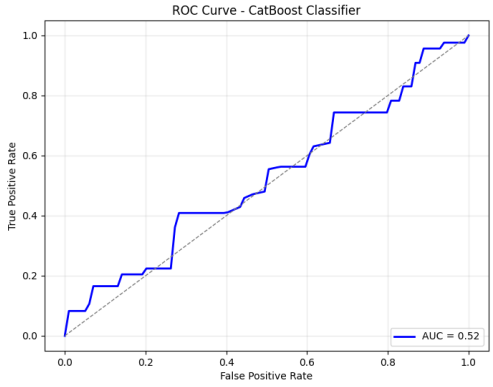


(a) ROC curve

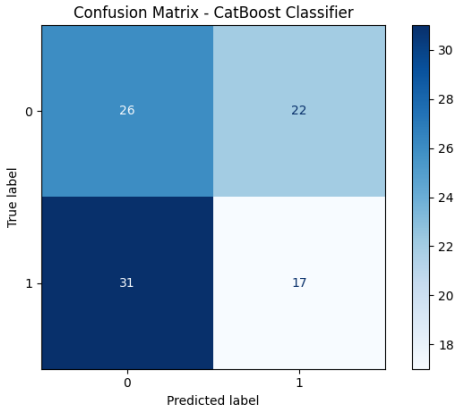


(b) Confusion matrix

Figure M.2: Evaluation metrics for the Random Forest classifier: AUC = 0.51; F1 = 0.37; precision = 0.65; recall = 0.40.

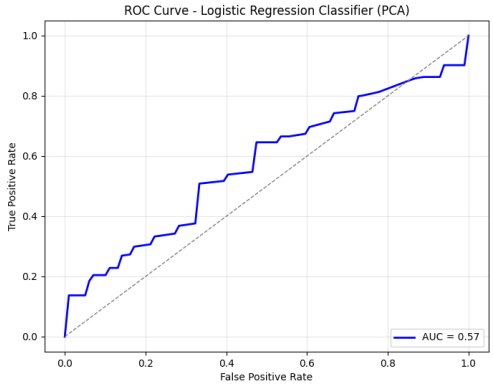


(a) ROC curve

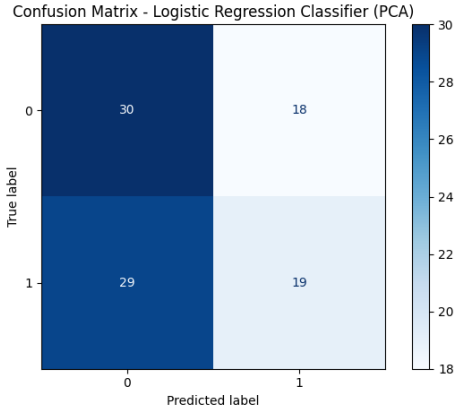


(b) Confusion matrix

Figure M.3: Evaluation metrics for the CatBoost Gradient Boosting Decision Tree classifier: AUC = 0.52; F1 = 0.37; precision = 0.53; recall = 0.37.

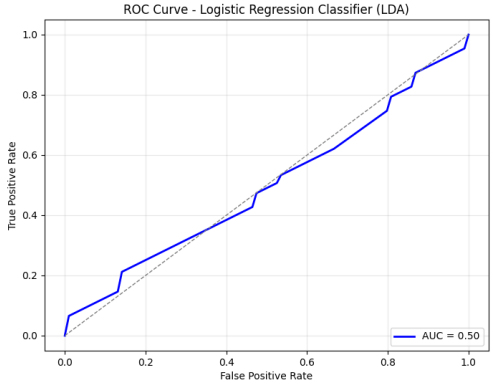


(a) ROC curve

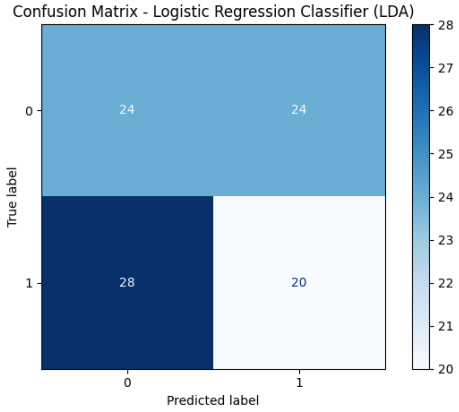


(b) Confusion matrix

Figure M.4: Evaluation metrics for the Logistic Regression classifier with PCA: AUC = 0.57; F1 = 0.42; precision = 0.65; recall = 0.41.

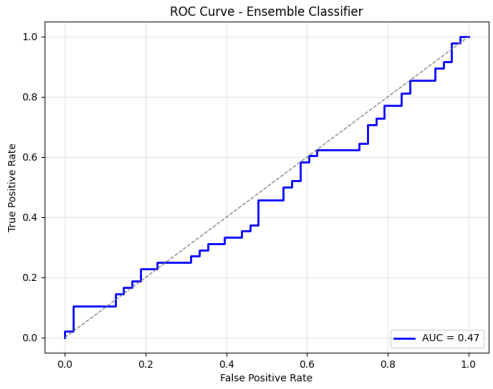


(a) ROC curve

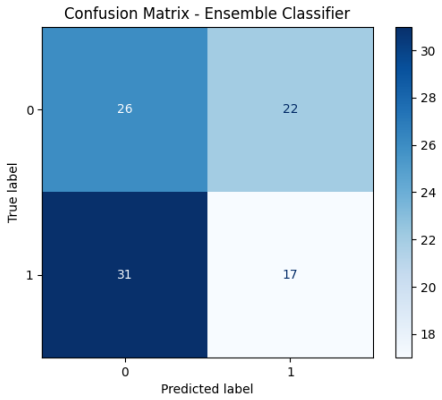


(b) Confusion matrix

Figure M.5: Evaluation metrics for the Logistic Regression classifier with LDA: AUC = 0.50; F1 = 0.34; precision = 0.31; recall = 0.45.



(a) ROC curve



(b) Confusion matrix

Figure M.6: Evaluation metrics for the Ensemble classifier averaging the Logistic Regression classifier, the Random Forest classifier and the CatBoost Gradient Boosting Decision Tree classifier: AUC = 0.47; F1 = 0.39; precision = 0.44; recall = 0.35.