2024

# GATEWAY TO THE MIND: USING EYE TRACKING TO EXPLORE THE BROADER AUTISM PHENOTYPE

Joseph Molski
*University of Rhode Island*, joseph_molski@uri.edu

# GATEWAY TO THE MIND: USING EYE TRACKING TO EXPLORE THE BROADER AUTISM PHENOTYPE

BY

JOSEPH MOLSKI

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

SPEECH-LANGUAGE PATHOLOGY

UNIVERSITY OF RHODE ISLAND

2024

MASTER OF SCIENCE THESIS

OF

JOSEPH MOLSKI

APPROVED:

Thesis Committee:

Major Professor Alisa Baron,

Vanessa Harwood

Paige Ramsdell

Brenton DeBoef
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2024

## ABSTRACT

The Broad Autism Phenotype (BAP) is a term used to describe Autism-like symptoms in individuals not diagnosed with autism spectrum disorder (ASD) as well as the genetic transmission of these symptoms. Within this field of study, first-degree relatives, siblings, and parents are most often studied due to the direct path of genetic transmission. Within this group, there are several measurable characteristics used to explore the BAP. Of relevance to this study are neuropsychological characteristics, specifically potential atypicality while gazing at speaking human faces. While some studies have conducted analyses with this neuropsychological characteristic with siblings, there has yet to be a study analyzing parents of children with ASD (pASD). To analyze potential atypicality in viewing a speaking face, the current study utilized eye tracking to measure differences in how pASD view a speaking face. Specifically, the experimental stimuli used was of a speaking face producing the consonant-vowel syllable /ba/. To assess how the participants viewed the speaking face, interest areas were created to track the duration of time spent looking at the speaking face across trials. Using linear mixed effects models, the results of this study indicated no significant differences between pASD and their neurotypical peers within this audiovisual paradigm. Future research with a larger sample size is necessary to validate these findings. In addition to a larger, more diverse population, future research may involve both parents of a child with ASD or other first-degree relatives, such as siblings.

## ACKNOWLEDGMENTS

I would like to express my deepest appreciation to all those who have supported me throughout this journey. From kind words to guidance and collaboration, I never felt alone in this process. This accomplishment would not have been possible without you.

First and foremost, I want to express my deepest appreciation and gratitude to my mentor and major professor, Dr. Alisa Baron. The wealth of opportunities, knowledge, and guidance you have graciously provided me over the past 3 years is astounding. Working alongside you has not only shaped my journey through higher education but has had an immense influence on who I am as a professional and as a person. You challenged me to think analytically, write clearly, and research effectively but most importantly to see the potential in myself. The lessons and experiences you have provided will remain with me for the rest of my life.

I would like to thank Dr. Harwood for not only committing your time to join my thesis committee but also for serving as a steadfast guide throughout my academic journey. Your empathy knows no bounds and your feedback, guidance, and support have been invaluable in my growth as a researcher and a clinician. I also would like to acknowledge and thank Dr. Paige Ramsdell for joining my thesis committee and dedicating your time to help refine this study.

Additionally, I would like to thank everyone at the University of Rhode Island who contributed to this study. To the faculty and staff of the Communicative Disorders Department for providing me with the skills and the tools I needed to complete this thesis. To my graduate cohort for your continued enthusiasm and encouragement

throughout this process, your words and actions have motivated me more than you know. To the student researchers at the Collaborative Cognitive Neuroscience Lab, you played an integral role in this study. Your meticulous work in collecting data and finding participants was invaluable. To Dr. Daniel Kleinman for your assistance in statistically analyzing the data and creating beautiful figures. Lastly, to all the participants who selflessly gave their time to this study. Your dedication allowed us to explore uncharted territory and contribute to this rapidly expanding area of research.

In closing, I would like to sincerely thank my friends and family. To the Molski and Turley families who supported me in innumerable ways and always checked in to see how I was doing. To my partner, Brielle, you have my deepest appreciation and admiration. You always stood by my side, and your presence kept me motivated and grounded throughout this process. You ceaselessly found new ways to support me and I am beyond grateful to have you in my life. To my sister and lifelong role model Kelly, words cannot express how profoundly you have impacted my life, I surely wouldn't be where I am today without you. You have inspired and motivated me my entire life. Finally, to my parents who made countless sacrifices to support me and provide me with amazing opportunities throughout my life. Without you, none of this would have been possible. You have opened doors for me that you couldn't access yourselves, and for that, I am eternally grateful.

## TABLE OF CONTENTS

## LIST OF TABLES

TABLE PAGE

## LIST OF FIGURES

FIGURE PAGE

# CHAPTER 1

## INTRODUCTION

The genetic component of autism spectrum disorder (ASD) has been a target for research for decades. In 1977, Folstein and Rutter published Infantile Autism: A Genetic Study of 21 Twin Pairs, the first publication comparing first-degree relatives in search of a genetic link involving ASD. Nearly fifty years later, researchers still seek to explore and define the genetic aspect of ASD, with Sasson et al. (2013) identifying a current need to examine the implication of genetics on the transmission of ASD and Tick et al. (2016) conducting a meta-analysis of twin studies regarding the heritability of ASD which ranged from 53% (in dizygotic twins) to 98% (in monozygotic twins). However, transmission represents only one aspect of the complex genetic picture of ASD, which leads us to the Broad Autism Phenotype (BAP). The BAP encompasses not only transmission but the genetic basis of ASD as a whole, including familial patterns, intergenerational transmission, and several observable endophenotypes (Persico et al., 2014).

Despite fifty years of research, there are still noticeable gaps in the BAP literature. One such gap involves the comparison of parents of children with Autism (pASD) to their neurotypical peers, which is the focus of this study. First-degree relatives have been the focus of BAP research before. The study of first-degree relatives not only allows for a comparison between genetically related individuals, but it also allows for the comparison of first-degree relatives to their neurotypical peers. Notable studies in this area include Eyuboglu et al. (2018) and

Petalas et al. (2013) who focused on siblings, as well as the aforementioned twin study by Folstein and Rutter. Through studying this population, common differences between first-degree relatives of people with ASD and their neurotypical peers can be identified and studied.

Common differences identified between first-degree relatives of people with ASD are often grouped and characterized. One such characterization was completed by Persico et al. (2014) which identified seven "Autism Endophenotypes", including; biochemical, morphological, hormonal, immunological, neurophysiological/neuroanatomical, neuropsychological, and behavioral. These endophenotypes classify and group observable and measurable "common differences" in first-degree relatives of people with ASD. Of relevance to this study is the neuropsychological endophenotype, which includes the "common difference" of abnormalities in visually scanning a human face.

To detect abnormalities in visually scanning a human face, specific tools must be used. In this study, eye tracking was utilized to track the gaze of participants while they attended to an audiovisual stimulus of a speaking face. Eye tracking permits researchers to record and analyze the way in which a participant visually engages with a stimulus. An audiovisual stimulus refers to an auditory stimulus presented in tandem with a visual stimulus, which in the case of this study is a video of a face producing single-syllable utterances. This creates a pseudo-communicative stimulus, as the stimulus does not communicate meaningful information. Eye tracking has been used before in similar studies, with Falck-Ytter et

al. (2013) conducting a methodological review of the use of eye tracking as a means of exploring ASD in children.

# CHAPTER 2

# REVIEW OF LITERATURE

## *The Broad Autism Phenotype*

Researchers have revealed the existence of a broader autism phenotype (BAP) in family members of individuals with ASD. According to Persico et al. (2014), there are at least seven categories where autism endophenotypes can be grouped; biochemical, morphological, hormonal, immunological, neurophysiological/neuroanatomical, neuropsychological, and behavioral. These endophenotypes can be used to trace the underlying genetic causation of ASD. Of these categories, we aim to explore the neuropsychological category, which is, in part, characterized by abnormalities in visually scanning a human face.

## *Identifying the Broad Autism Phenotype*

While several studies have investigated the communication and cognitive abilities in pASD, there has yet to be an investigation that has measured gaze patterns to the face during speech production by this group. While there are documented differences in the way that individuals with ASD scan and process a face compared to their neurotypical peers (Lai et al., 2013), this has yet to be explored in pASD. By analyzing these differences, we aim to examine the similarities and differences in gaze patterns between pASD and the neurotypical control group to understand the

neuropsychological characteristics associated with the BAP within an audiovisual integration paradigm.

As previously stated, no evaluation of the BAP has been conducted for parents of children with ASD. However, researchers have conducted studies on the interaction of BAP and siblings, concluding that siblings of children with ASD have autistic-like traits (Eyuboglu et al., 2018; Petalas et al., 2013). Additionally, research on siblings of children with ASD has been conducted using similar methodologies to the current study, such as Dalton et al. (2006), who utilized functional magnetic resonance imaging (fMRI) in conjunction with eye tracking. This study found that siblings of individuals with ASD displayed notable disparities in gaze fixation and brain activation patterns when observing static images of human faces when compared to typically-developing peers without ASD-affected siblings. This demonstrated that eye tracking can be used to effectively investigate first-degree relatives within the BAP. Thus, the use of eye-tracking technology in the current study with audio and video stimuli of a speaking face may be a salient method for identifying information to contribute to the understanding of the BAP.

*Eye Tracking*

Researchers have documented those individuals with ASD scan faces differently from their neurotypical peers (Falck-Ytter et al., 2013). Children diagnosed with ASD had reduced visual attention to the face of a speaker and are less fixated on the speaker's mouth than typically developing individuals when presented with audiovisual stimuli (Irwin & Brancazio, 2014). Additionally, in a review of

communication abilities within the BAP, Gerdts and Bernier (2011) stated that the most common finding in undiagnosed family members of individuals with ASD were mild impairments of social and communication capabilities that are akin to the capabilities of individuals with ASD but to a smaller extent. This may indicate that undiagnosed family members of individuals with ASD, such as parents, will demonstrate a difference in scanning a face during an audiovisual task similar to children diagnosed with ASD.

This study uses an audiovisual integration paradigm. This type of stimulus is novel to research on the BAP. Dynamic speaking faces may be particularly relevant to social communication due to the integration of visual information from the face and auditory information from spoken language. In previous studies, researchers have analyzed gaze patterns on static faces with no auditory stimuli, which does not reflect typical communication (Chawarska et al., 2009; Dalton et al., 2017; de Wit et al., 2008). The current study's experimental design uses a stimulus that is more ecologically valid than related studies, a result of data collection involving dynamic rather than static facial stimuli during the audiovisual integration task. This design allows for a greater exploration of potential impairments of social communication abilities in undiagnosed first-degree relatives as a dynamic face is more comparable to social communication than a static face.

*Audiovisual Integration*

Audiovisual (AV) integration occurs when an individual consolidates an auditory stimulus and a visual stimulus. This is a documented phenomenon,

especially in the context of spoken language being influenced by visual information known as phonemic restoration (e.g., Irwin et al., 2022; 2023; Warren, 1970). Numerous researchers have investigated audiovisual integration and explored its effects during speech across a range of different populations and in a variety of different conditions (Grant et al., 1998; Sommers et al., 2005; Sumby and Pollack, 1954). AV integration often occurs during speech in the context of face-to-face communication, as it generally involves the consolidation of auditory and visual elements to receive and interpret a message.

Studies comparing typically-developing individuals and individuals with ASD on AV integration tasks suggest that there is a difference in the way these groups integrate the aforementioned stimuli. Ronconi et al. (2023) found that a "disproportionate elaboration of the auditory input could be the main factor characterizing atypical audiovisual integration in autism". In addition to this, Brandwein et al. (2013) found that during an AV integration task, typically-developing children demonstrated behavioral support for multisensory (auditory and visual) input, while children diagnosed with ASD did not. Despite this research into AV integration for individuals with ASD, there has yet to be an investigation of this phenomenon with pASD compared to their neurotypical peers. To address this gap in the literature, the current study aims to use eye tracking during an AV integration task to better determine differences between parents of children with ASD and their neurotypical peers concerning gaze patterns on a speaking face. This may provide valuable information on the neuropsychological aspects of the BAP.

While there are several methods to evaluate AV integration, the current study includes an experiment that utilizes pixelation as a means of removing meaningful visual information from the mouth during an AV task. By removing the meaningful elements of a visual stimulus, the impact of integration can be studied. This experimental task was done through a phonemic restoration paradigm. Two auditory stimuli, /ba/ and /a/, were presented within the two experimental conditions, audiovisual and pixelated. Furthermore, there were two experiments regarding participant responses, active when participants were asked to respond and passive when participants were asked to view the stimuli passively. It is expected that in the audiovisual condition, when the audio is /a/, the presence of the speaking face producing the consonant and vowel /ba/ will "repair" the consonant. This means that when participants hear /a/ they will perceive /ba/. In the pixelated condition, it is expected that this "repair" will not occur as visual information from the mouth is not present.

This led to our research question: Can gaze patterns measured through eye tracking be used to identify differences between parents of children with ASD and their neurotypical peers?

# CHAPTER 3

# METHODOLOGY

## *Participants*

This study recruited neurotypical adults who were native English speakers residing in Rhode Island. To be eligible for participation, all participants needed to meet specific criteria, including being monolingual English speakers without any known neurological impairments such as concussions, seizures, epilepsy, traumatic brain injury, or neurodegenerative disorders. Audiological and visual screenings were conducted since this study involved the presentation of both audio and visual stimuli. For the audiometry screening, a portable Grason-Stadler GSI 18 screening audiometer and headphones were used to test both ears at frequencies ranging from 500-4K Hz, with a threshold of 25 dB. All participants successfully passed this screening. Additionally, participants received a vision acuity screening using a Snellen Eye Chart with multiple letters. The participants stood 20 feet away from the chart displayed on a wall and were required to read line 8 with both eyes open (aided with glasses if needed). This line corresponded to 20/40 vision, and all participants passed the vision acuity screening.

A total of fifteen participants who met these criteria were included in the study, consisting of 11 females and 4 males with a mean age of 43.14 ($SD$ = 11.70). To provide a further description of the participant group, their race/ethnicity information was also recorded. Out of the participants, 12 identified as White, 1 as Black/African

American, 1 as Asian, and 1 did not report their race/ethnicity. The pASD group consisted of 6 females and 2 males with a mean age of 44.38 years old (*SD* = 11.72), race/ethnicity is as follows: 7 identified as White, 1 identified as African American. The neurotypical group consisted of 5 females and 2 males with a mean age of 41.74 years old (*SD* = 13.33), race/ethnicity is as follows: 5 identified as white, 1 identified as Asian, and 1 did not report their race/ethnicity. These groups were gender and age-matched with a +/- 2 year age difference with 0.90 SD.

Participants underwent a cognitive assessment to evaluate their nonverbal IQ. The cognitive test was conducted either on the same day as the audiovisual integration experiment or during a separate session within a week, ensuring that all testing occurred within a short timeframe. The Weschler Abbreviated Scale of Intelligence-Second Edition (WASI-2; Wechsler, 2011) was used to assess cognitive abilities. Specifically, the Block Design and Matrix Reasoning subtests of the WASI-2 were administered, which contributed to the Perceptual Reasoning composite score. The Perceptual Reasoning composite measures non-verbal reasoning. It is a standard score with a mean of 100 and a standard deviation of 15; therefore, the average range lies between 85-115. The participants obtained the following scores on the Perceptual Reasoning composite of the WASI-2: pASD: mean = 92.5 (*SD* = 13.47, range = 75-112) with 2 participants below the average range, for neurotypical peers: mean = 100.71 (*SD* = 8.32, range = 93-113).

*Procedure*

The participants participated in either 1 or 2 sessions depending on availability, lasting around 1.5 hours each. In these sessions, they underwent experimental procedures involving both eye tracking and EEG, as well as behavioral assessments. The data used for this thesis was collected within a larger study on phonemic restoration, which included an EEG task. During this research, the EEG and eye-tracking data were collected in tandem. However, since the main research question is best answered using eye-tracking data, EEG data will not be discussed here.

*Eye-tracking stimuli*

In this study, the stimuli used are identical to those employed in the studies conducted by Irwin et al. (2017) and Baron et al. (2023). The stimuli were generated by recording a male adult speaker who produced the English syllable sound /ba/ in a soundproof room. A digital video camera positioned approximately 3 feet away from the speaker captured the visual recording. Synthetic speech was then generated based on the natural production of the syllable, however, specific modifications were made to flatten the formant transitions for the /a/ syllable. The video stimulus of the speaker's face remained unaltered and consistently displayed the face during the production of the /ba/ visual stimulus. However, the auditory stimuli varied between /ba/ and /a/, resulting in the visual cues compensating for the weakened auditory cues. This phenomenon occurs when the auditory stimulus /a/ is presented alongside the visual stimulus /ba/, causing the perceived stimulus to be interpreted as /ba/. This

effect is similar to the visual phonemic restoration effect observed in previous studies (Irwin et al., 2011; 2014; 2017; 2021; Kashino, 2006; Samuel, 1981).

Separate high-quality audio was recorded using Praat (Boersma & Weenink, 2016) and a Sennheiser microphone on a Macbook Pro. The microphone was positioned centrally and was approximately 2 feet posterior to the video camera and in front of the speaker. The /ba/ tokens were repeated five times for five iterations, resulting in a total of 25 total tokens. One token served as the basis for all stimuli. Using Praat, various acoustic parameters were extracted from this token, including formant trajectories, amplitude contour, voicing, and pitch contour. The token exhibited rising formant transitions for F1, F2, and to a lesser extent F3, which are characteristic of /ba/. To generate the /ba/ stimulus, a novel token of /ba/ was synthesized based on these values. For the /a/ stimulus, the synthesis parameters were modified. As a result, the onset values for F1 and F2 were modified to reduce the extent of the transitions, and the transition durations for F1, F2, and F3 were extended. Subsequently, a new stimulus was synthesized. In the case of /ba/, the transitions lasted for 34 ms, with F1 rising from 500 Hz to 850 Hz, F2 rising from 1150 Hz to 1350 Hz, and F3 rising from 2300 Hz to 2400 Hz. For the /a/ stimulus, the transitions lasted for 70 ms, with F1 rising from 750 Hz to 850 Hz, F2 rising from 1300 Hz to 1350 Hz, and F3 rising from 2300 Hz to 2400 Hz (refer to Figure 1 for spectrograms of /ba/ and /a/).

As the audio was recorded simultaneously but separately from the video, the synthesized auditory stimuli for /ba/ and /a/ were dubbed onto a video of the speaker producing /ba/ (audiovisual condition). The acoustic onsets were synchronized with

the visible articulation in this condition. Alternatively, the stimuli were dubbed onto a video of a face with a pixelated (PX) mouth region that exhibited no visible movement (PX condition). In the PX condition, the mouth portion of the video was pixelated into 36 48x48 solid blocks. The mouth region itself was confined within 9 of these blocks, forming a 3x3 grid. This pixelation ensured that the articulatory movements of the mouth and jaw were not perceivable, although changes in pixelation indicated movement. All stimulus videos can be accessed publicly at https://osf.io/ehvg8/.

*Eye-tracking procedure*

The study took place in a room without windows. Before testing, each participant participated in a calibration and validation process for the eye tracker, consisting of five points on the screen. If any x- or y-axis drift exceeding 2° was detected, recalibration and validation were performed. In general, the majority of participants experienced drift below 1.25°. Participants were explicitly instructed to keep their gaze fixed on the screen, as looking away from the screen would cause the experiment to pause.

The experiment was designed within Experiment Builder software (version 1.10.165), developed by SR Research. The stimuli were presented in both audiovisual (AV) and pixelated (PX) conditions. Additionally, the stimuli were arranged and presented in a 70/30 oddball design, meaning that the /ba/ sound served as the frequently occurring standard stimulus (presented 70% of the time) while the /a/ sound served as the infrequently occurring deviant stimulus (presented 30% of the time), in both pixelation states. The first part of the experiment was passive, meaning no action

was required by the participant while the second part of the experiment was active, involving a button press. Again, the stimuli were presented in both AV and PX conditions. Participants always completed the passive experiment first, followed by the active experiment. In both experiments, the AV condition was presented first, followed by the PX condition. This order ensured that the phonemic restoration effect was examined without participants being exposed to the contrast between the /ba/ and /a/ auditory tokens before the experiment. In the active experiment, participants were instructed to press a button on a gamepad to indicate whether they heard a standard or deviant stimulus. At the beginning of the active experiment and before each condition, participants listened to the /a/ and /ba/ sounds and were informed which button (right or left) corresponded to each sound. This reminder was provided to participants because the buttons on the gamepad were not labeled, ensuring that they did not have to look away from the monitor. Each condition (AV and PX) lasted for 9 minutes and included 200 trials (140 standard and 60 deviant), with each trial lasting 2000 ms. The total duration of the experiment, including all four blocks/conditions, was 36 minutes. The time between trials, known as the interstimulus interval (ISI), was set to 200 ms. The audio was adjusted to a comfortable listening level, approximately 65 dB. Before each block, a drift check was performed for eye-tracking accuracy. Additionally, there were five practice trials at the beginning of each AV and PX condition for both the passive and active experiments.

*Apparatus*

Data was collected at the University of Rhode Island with an EyeLink Portable Duo. Monocular data was collected with a sampling rate of 500 Hz. Sentence stimuli were displayed on a 19-inch PC monitor. The display resolution was presented as 1280 x 1024. The viewing distance of the seated participants was between 80 and 90 cm. EEG data and eye-tracking data were obtained simultaneously, thus, the eye tracker remote mode was utilized and a target sticker was positioned on the EEG cap which was then placed on the participants' forehead, which allowed for the system to compensate for head movements of a distance up to 20 cm.

*Data Preprocessing*

To preprocess the data, specific facial regions were identified as areas of interest. These areas included the eyes, nose, mouth/jaw, and the head. The EyeLink Data Viewer (version 4.1.1) was used to define these regions on the target face. The head interest area encompassed the entire target head, including the ears and hair. The eye interest area was defined as a rectangular region extending from the upper part of the brow to the top of the cheekbone. It is common in the literature to combine data from both eyes in analyses (e.g., Falck-Ytter, 2008). The mouth/jaw interest area was manually drawn and connected to the nose, including the philtrum, and extending along the base of the jaw and laterally to the nasolabial sulcus. It is important to note that the mouth/jaw interest area was designed to encompass the entire mouth, including its open posture and movement during test trials. The nose region was defined by the lower part of the eye interest area and the upper region of the mouth/jaw interest area, extending laterally to cover the width of the nose.

*Data Cleaning Procedure*

The eye-tracking data was exported using EyeLink Data Viewer software, and the data cleaning process consisted of multiple stages. The data was then subjected to a three-stage cleaning process. First, fixations lasting less than 80 ms and within 0.5° were combined with nearby fixations. Second, fixations lasting less than 40 ms and within 1.25° were merged with neighboring fixations. Finally, any remaining fixations shorter than 80 ms were excluded from the analysis. Previous research has indicated that a minimum duration of 50 ms is required for the eye-to-brain lag, ensuring the extraction of useful visual information for processing (Inhoff & Radach, 1998). Short fixations are generally considered to be microsaccades, truncated fixations, blinks, or other artifacts in eye movement data (Godfroid, 2020). Therefore, merging or removing short fixations is recommended, as they are not believed to reflect cognitive processing (Godfroid, 2020). Following the three-stage cleaning process, trials were discarded if the participant had no fixations on the face (1.06%).

*Analysis*

Initially, using Data Viewer, heat maps of the speaking face were created in order to visualize the eye-tracking data between pASD and their neurotypical peers. Then, fixations were analyzed over the course of the trials (to provide a dynamic measure of differences in looking patterns to articulators as the speaking face produces the /ba/ or /a/. A Time Course (Binning) report was used to export the data and binned

into 20 ms bins. Samples that fall outside of the predefined interest areas during saccades and blinks were excluded. All subsequent analyses were conducted in R.

Prior to analysis, fixations to each interest area were averaged within four consecutive, non-overlapping 300 ms bins (consistent with previous work, Baron et al., 2023; Irwin & Brancazio, 2014). The resulting bins appropriately segment the visual and auditory information corresponding to the initial rest position of the face (0-300 ms), the opening of the mouth before the closing gesture of the consonant (300-600 ms), the closing gesture of the consonant (600-900 ms), and the peak mouth opening for the vowel 900-1200 ms).

Using the *lme4* package (v.1.1-21; Bates et al., 2015) trial-level analyses were performed using linear mixed-effects models (LMEM), with a single model paired to each of the interest areas (Eyes and Mouth/Jaw). For this model, the dependent variable was the proportion of fixations on every trial, in each time window for that specific interest area. All models had fixed effects of experiment (two levels: passive = -0.5, active = +0.5), condition (two levels: pixelated (PX) = -0.5, audiovisual (AV) = +0.5), group (two levels: pASD = -0.5, neurotypical peers = +0.5), time window (a continuous, centered variable, with contrast weights -1.5, -0.5, +0.5 and +1.5 so that adjacent time windows were separated by 1), stimulus (two levels: standard /ba/ = -.3, deviant /a/ = +0.7), and within-task trial number (centered and scaled to have a range of 1). All interactions between these fixed effects were included, and the maximal random effects structure was supported by the data. To identify the maximal random effects structure, we followed a three-step procedure. Initially, we used the *bobyqa* optimizer to fit a model with a maximal random effects structure. This included a

random intercept for participants, all within-factor random slopes and their interactions, and correlations between random slopes. If the model did not converge, we removed correlations between random slopes. If the resulting model still did not converge, we identified random slopes accounting for <0.01% of the variance of their associated random factors and then removed all such slopes at the same time (Bates et al., 2018). Main effects and interactions were evaluated with the *contestMD* function from the *lmerTest* package. For significant effects, follow-up contrasts were applied to the fitted model using the *emmeans* package (v. 1.7.1-1; Lenth, 2021). To account for two models (one for each interest area), a Bonferroni correction for two comparisons (adjusted alpha = 0.025) was used for all tests. The Tukey method was used to control for the familywise error rate when conducting pairwise comparisons.

# CHAPTER 4

# FINDINGS

We plotted grand mean proportions of fixations to 5 interest areas (mouth/jaw, eyes, face, and nose) over the course of each trial for each experiment (active and passive) and condition (AV and PX) for the neurotypical (Figure 1) and pASD groups (Figure 2). Ninety-five percent confidence intervals are shown as error ribbons while the dashed vertical lines indicate the four time windows (at 300 ms intervals). Figure 3 shows a visual representation of both groups (pASD and neurotypical) for only the mouth/jaw and eye areas of interest. Throughout the experiment, as the audiovisual stimuli unfolded, the participants' gaze shifted toward the mouth during un-pixelated trials, especially during active trials, which required a participant's response.

Linear mixed-effects models (LMEM) were used to analyze the eye movements of pASD and their neurotypical peers within two pre-specified regions (eyes and mouth/jaw) when viewing a speaking face. As previously mentioned, the dependent variable was the proportion of fixations on every trial, in each time window for that specific region of interest (eyes or mouth/jaw).

All statistical tests are reported in Tables 1 through 4. To address our research question regarding the gaze patterns to a speaking face we examined the mouth/jaw and eye areas. In general, across all participants, there were more looks to the mouth region throughout the trial (Table 1). There were more looks to the mouth in the active experiment than in the passive experiment and there were also more looks to the

mouth in the audiovisual condition than in the pixelated condition. Over the course of the trial, mouth fixations significantly increased more during the active experiment than the passive experiment. Additionally, throughout the trial, mouth fixations significantly increased more during the audiovisual condition than the pixelated condition. The effect of the condition (audiovisual vs. pixelated) was stronger for the active experiment than the passive. Furthermore, throughout the trial, mouth fixations were significantly different between the audiovisual and pixelated conditions for the active experiment but not for the passive experiment. Regarding the group effects of the mouth region, no significant differences were noted (Table 3).

In general, across all participants, there were fewer looks to the eyes over the course of the trial (Table 2). Over the course of the trial, eye fixations significantly decreased more during the active experiment than the passive experiment. Furthermore, throughout the trial, eye fixations were significantly different between the audiovisual and pixelated conditions. Regarding the group effects of the mouth region, no significant differences were noted (Table 4).

Figure 1.

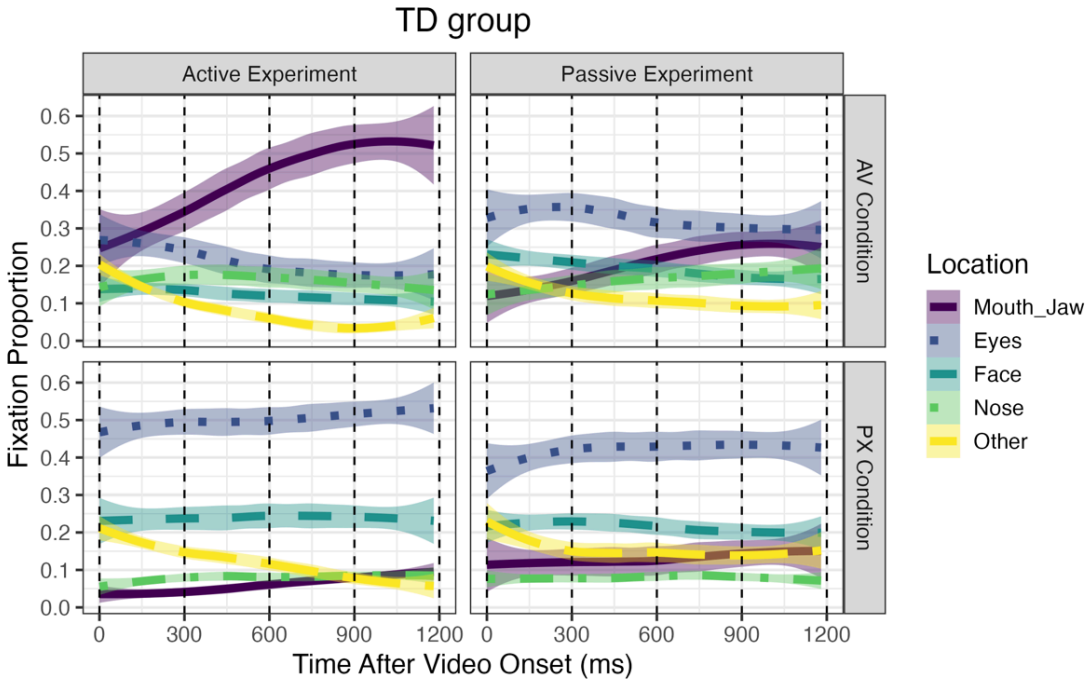


Figure 2.

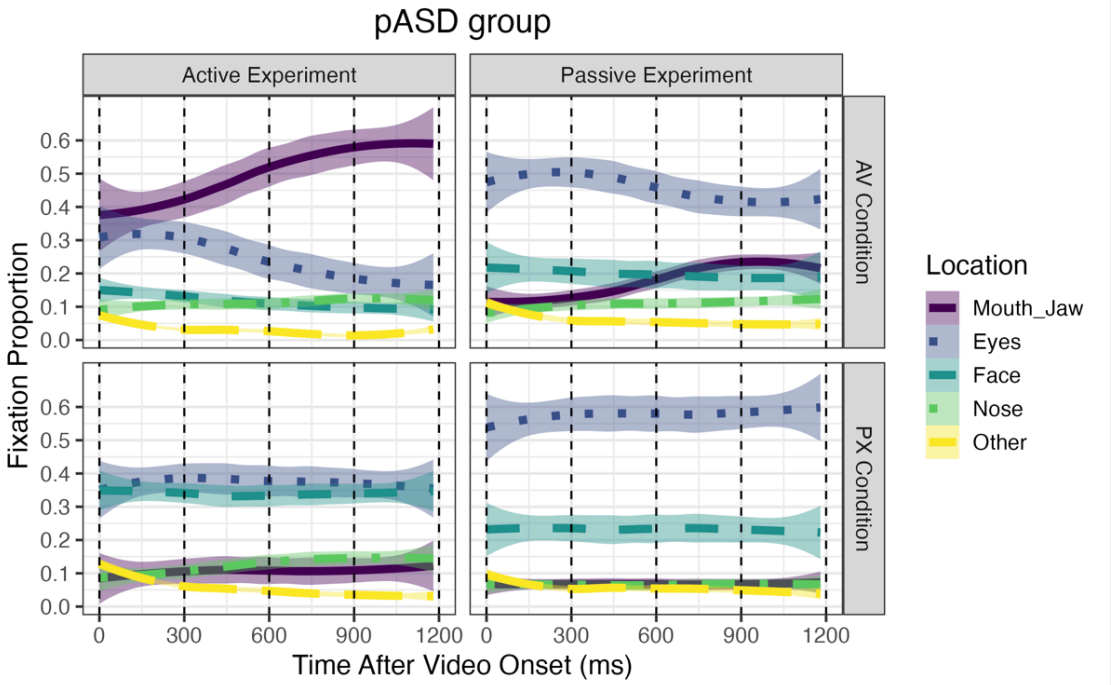

Figure 3.

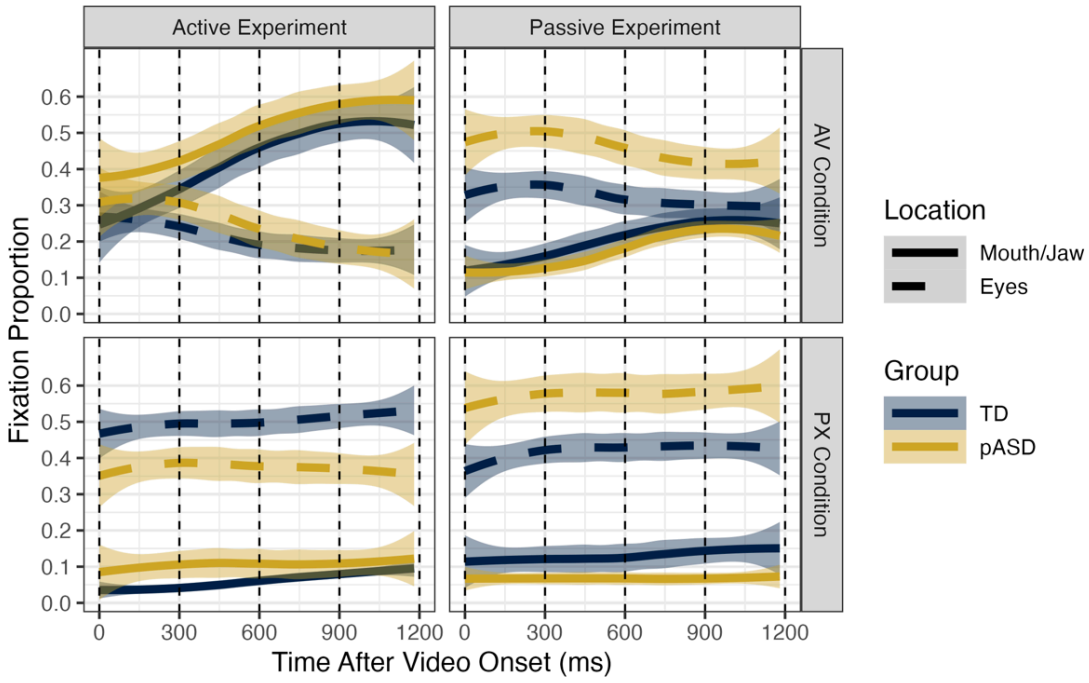
Active Experiment
Passive Experiment
AV Condition
PX Condition
Fixation Proportion
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
300
600
900
1200
Time After Video Onset (ms)
Location
Mouth/Jaw
Eyes
Group
TD
pASD

Table 1. **Eye Region to Mouth/Jaw Region Comparison**

| Region | Effects | β | Std. Error | df | *t* | *p* |
|---|---|---|---|---|---|---|
| Mouth/Jaw | Intercept | 0.208 | 0.041 | 13.202 | 5.005 | <.001 |
| | Across time windows | 0.033 | 0.006 | 13.184 | 5.31 | <.001 |
| | Active - Passive | 0.126 | 0.039 | 12.530 | 3.229 | .006 |
| | Audiovisual - Pixelated | 0.235 | 0.043 | 13.073 | 5.422 | <.001 |
| | Increase over time windows across Active - Passive | 0.019 | 0.007 | 12.707 | 2.487 | .027 |
| | Increase over time windows across Audiovisual - Pixelated | 0.049 | 0.009 | 13.118 | 5.218 | <.001 |
| | Active – Passive * Audiovisual - Pixelated | 0.288 | 0.098 | 13.206 | 2.926 | .011 |
| | Increase over time windows across Active – Passive * Audiovisual - Pixelated | 0.0279 | 0.0153 | 12.867 | 1.822 | .091 |

Table 2. **Eye Region to Mouth/Jaw Region Comparison**

| Region | Effects | β | Std. Error | df | *t* | *p* |
|---|---|---|---|---|---|---|
| Eyes | Intercept | 0.394 | 0.05 | 12.758 | 7.765 | 3.48E-06 |
| | Across time windows | -0.012 | 0.004 | 12.957 | -2.956 | .011 |
| | Active - Passive | -0.103 | 0.053 | 12.161 | -1.924 | .078 |
| | Audiovisual - Pixelated | -0.160 | 0.058 | 13.240 | -2.762 | .015 |
| | Increase over time windows across Active - Passive | -0.010 | 0.01 | 13.227 | -0.982 | .343 |
| | Increase over time windows across Audiovisual - Pixelated | -0.038 | 0.009 | 13.089 | -3.914 | .001 |
| | Active – Passive * Audiovisual - Pixelated | -0.103 | 0.090 | 12.976 | -1.146 | .272 |
| | Increase over time windows across Active – Passive * Audiovisual - Pixelated | -0.009 | 0.015 | 12.885 | -0.615 | .549 |

Table 3. **Group Effects (Mouth/Jaw)**

| Region | Effects | β | Std. Error | df | *t* | *p* |
|---|---|---|---|---|---|---|
| Mouth/Jaw | Group | 0.004 | 0.083 | 13.202 | 0.059 | .953 |
| | Group*Across time windows | -0.009 | 0.012 | 13.184 | -0.792 | .442 |
| | Group*Active - Passive | 0.1 | 0.078 | 12.53 | 1.274 | .225 |
| | Group* Audiovisual - Pixelated | 0.03 | 0.086 | 13.073 | 0.35 | .732 |
| | Group*Increase over time windows across Active - Passive | -0.007 | 0.015 | 12.707 | -0.505 | .621 |
| | Group*Increase over time windows across Audiovisual - Pixelated | 0.001 | 0.019 | 13.118 | 0.102 | .92 |
| | Group*Active – Passive * Audiovisual - Pixelated | -0.01 | 0.197 | 13.206 | -0.054 | .957 |
| | Group*Increase over time windows across Active – Passive * Audiovisual - Pixelated | -0.011 | 0.03 | 12.867 | -0.377 | .712 |

Table 4. **Group Effects (Eye)**

| Region | Effects | β | Std. Error | df | *t* | *p* |
|---|---|---|---|---|---|---|
| Eye | Group | 0.059 | 0.101 | 12.758 | 0.582 | .57 |
| | Group*Across time windows | -0.012 | 0.008 | 12.957 | -1.512 | .154 |
| | Group*Active - Passive | -0.171 | 0.107 | 12.161 | -1.601 | .134 |
| | Group*Audiovisual - Pixelated | 0.07 | 0.116 | 13.24 | 0.61 | .552 |
| | Group*Increase over time windows across Active - Passive | -0.012 | 0.021 | 13.227 | -0.595 | .562 |
| | Group*Increase over time windows across Audiovisual - Pixelated | -0.005 | 0.019 | 13.089 | -0.289 | .777 |
| | Group*Active – Passive * Audiovisual - Pixelated | 0.184 | 0.180 | 12.976 | 1.021 | .326 |
| | Group*Increase over time windows across Active – Passive * Audiovisual - Pixelated | 0.005 | 0.031 | 12.885 | 0.164 | .872 |

# CHAPTER 5

# CONCLUSION

## Discussion

The objective of this study was to compare pASD to their neurotypical peers using an audiovisual paradigm of a speaking face. It was hypothesized that when viewing a video stimulus of a speaking face, there would be measurable differences in the gaze patterns between groups. Additionally, it was hypothesized that audiovisual and pixelated conditions would highlight group differences by removing functional articulatory information from the mouth. These hypotheses were selected due to their potential to contribute information to a greater understanding of the BAP.

The results of the study proved the null hypothesis, indicating no significant differences between groups across both experiments and conditions. Previous studies have revealed that compared to their neurotypical peers, individuals with ASD scan and process a face differently (Lai et al., 2013) and that first-degree relatives of individuals with ASD have autistic-like traits (Dalton et al., 2006; Eyuboglu et al., 2018; Petalas et al., 2013). The current study found no significant differences between first-degree relatives of individuals with ASD and their neurotypical peers.

The results of this study indicate that gaze patterns measured through eye tracking did not indicate a difference between parents of children with ASD and their neurotypical peers. This conclusion disputes previous studies such as (Dalton et al. 2006) who used similar methodologies and found measurable differences in gaze

fixation and brain activation patterns of siblings of individuals with ASD when observing static images of human faces in comparison to typically-developing peers without ASD-affected siblings. Both studies observed first-degree relatives of individuals with ASD however the experiments differed. While Dalton et al. (2006) utilized static stimuli, the current study used audiovisual stimuli with active and passive experiments, which may explain the differing results.

**Limitations**

Difficulties were encountered in recruiting pASD for this study, with only eight participants matching the study's criteria. Potential explanations for this difficulty are the time dedication required to participate in this research which, as previously mentioned, lasted approximately 3 hours. pASD may have less time to commit as a result of caring for their child who due to the nature of this population, may have complex needs. Furthermore, the occurrence of the COVID-19 pandemic during the data collection period may have dissuaded individuals from committing to in-person research endeavors. Finally, the experimental criteria for this study allowed a minimum of one parent of a child with ASD to participate. This limits the ability to measure phenotypic characteristics of the experimental group, as both parents are contributing factors. Future research on this population may utilize childcare and provide support for pASD in order to improve recruitment efforts. Additionally, multiple data collection locations may be required to increase population size/diversity and reduce potential travel time for participants.

**Future Directions**

Due to the novel methodology and experimental design utilized in this study, there are several future directions that this research may take. Future studies may require both genetic parents of a child with ASD as criteria for participation in the study. As previously mentioned, the current study did not require both parents to participate, and by including both parents, additional information on the BAP may be investigated. An alternate direction that may be taken is to apply this hypothesis and experimental paradigm to a new population of first-degree relatives, such as siblings or even an entire familial unit of a person with ASD. Finally, developing sentence-length audiovisual stimuli instead of single consonants would be more ecologically valid as it would be more representative of typical communication.

BIBLIOGRAPHY

Altieri. (2013). Audiovisual integration: An introduction to behavioral and neuro-cognitive methods. *Frontiers in Psychology,* 4, 642–642. https://doi.org/10.3389/fpsyg.2013.00642

Baron, A., Harwood, V., Kleinman, D., Campanelli, L., Molski, J., Landi, N., & Irwin, J. (2023). Where on the face do we look during phonemic restoration: An eye-tracking study. *Frontiers in Psychology,* 14, 1005186–1005186. https://doi.org/10.3389/fpsyg.2023.1005186

Brandwein, A. B., Foxe, J. J., Butler, J. S., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2013). The development of multisensory integration in high-functioning autism: High-density electrical mapping and psychophysical measures reveal impairments in the processing of audiovisual inputs. *Cerebral Cortex, 23*(6), 1329–1341. https://doi.org/10.1093/cercor/bhs109

Chawarska, K., & Shic, F. (2009). Looking but not seeing: atypical visual scanning and recognition of faces in 2 and 4-year-old children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 39*(12), 1663–1672. https://doi.org/10.1007/s10803-009-0803-7

Dalton, K. M., Nacewicz, B. M., Alexander, A. L., & Davidson, R. J. (2007). Gaze-fixation, brain activation, and amygdala volume in unaffected siblings of individuals with autism. *Biological Psychiatry, 61*(4), 512–520. https://doi.org/10.1016/j.biopsych.2006.05.019

De Wit, T., Falck-Ytter, T., & von Hofsten, C. (2008). Young children with Autism Spectrum Disorder look differently at positive versus negative emotional faces. *Research in Autism Spectrum Disorders, 2*(4), 651–659. https://doi.org/10.1016/j.rasd.2008.01.004

Eyuboglu, M., Baykara, B., & Eyuboglu, D. (2018). Broad autism phenotype: Theory of mind and empathy skills in unaffected siblings of children with autism spectrum disorder. *Psychiatry and Clinical Psychopharmacology, 28*(1), 36–42. https://doi.org/10.1080/24750573.2017.1379714

Gerdts, J., & Bernier, R. (2011). The broader autism phenotype and its implications on the etiology and treatment of autism spectrum disorders. *Autism Research and Treatment,* 2011, 545901. https://doi.org/10.1155/2011/545901

Godfroid, A., & Winke, P. M. (2019). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. Routledge.

Grant, K., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *The Journal of the Acoustical Society of America, 103*(5 I), 2677–2690. https://doi.org/10.1121/1.422788

Gredeback, G., Bolte, S., & Falck-Ytter, T. (2013). Eye tracking in early autism research. *Journal of Neurodevelopmental Disorders, 5*(1), 28–28. https://doi.org/10.1186/1866-1955-5-28

Inhoff, A. W., &Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. *Eye Guidance in Reading and Scene Perception,* 29–53. https://doi.org/10.1016/b978-008043361-5/50003-1

Irwin, J. & Brancazio, L. (2014). Seeing to hear? Patterns of gaze to speaking faces in children with autism spectrum disorders. *Frontiers in Psychology, 5,* 397. https://doi.org/10.3389/fpsyg.2014.00397

Irwin, J., Avery, T., Kleinman, D., & Landi, N. (2022). Audiovisual speech perception in children with autism spectrum disorders: Evidence from visual phonemic restoration. *Journal of Autism and Developmental Disorders, 52*(1), 28–37. https://doi.org.uri.idm.oclc.org/10.1007/s10803-021-04916-x

Lai, M. C., Lombardo, M. V., Chakrabarti, B., & Baron-Cohen, S. (2013). Subgrouping the autism "spectrum": Reflections on DSM-5. *PLoS Biology, 11*(4), e1001544. https://doi.org/10.1371/journal.pbio.1001544

Losh, M., Childress, D., Lam, K., & Piven, J. (2008). Defining key features of the broad autism phenotype: A comparison across parents of multiple- and single-incidence autism families. *American Journal of Medical Genetics Part B, Neuropsychiatric Genetics, 147B*(4), 424–433. https://doi.org/10.1002/ajmg.b.30612

Nuske, H., Vivanti, G., & Dissanayake, C. (2016). Others' emotions teach, but not in autism: An eye-tracking pupillometry study. *Molecular Autism, 7*(1), 36–36. https://doi.org/10.1186/s13229-016-0098-4

Persico, A.M., Sacco, R. (2014). Endophenotypes in autism spectrum disorders. In: Patel, V., Preedy, V., Martin, C. (eds). Comprehensive Guide to Autism. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-4788-7_1

Petalas, M., Hastings, R. P., Nash, S., Hall, L. M., Joannidi, H., & Dowey, A. (2012). Psychological adjustment and sibling relationships in siblings of children with autism spectrum disorders: Environmental stressors and the broad autism phenotype. *Research in Autism Spectrum Disorders, 6*(1), 546–555. https://doi.org/10.1016/j.rasd.2011.07.015

Ronconi, L., Vitale, A., Federici, A., Mazzoni, N., Battaglini, L., Molteni, M., & Casartelli, L. (2023). Neural dynamics driving audio-visual integration in autism. *Cerebral Cortex , 33*(3), 543–556. https://doi.org/10.1093/cercor/bhac083

Sasson, N., Lam, K. S. L., Parlier, M., Daniels, J. L., & Piven, J. (2013). Autism and the broad autism phenotype: Familial patterns and intergenerational transmission. *Journal of Neurodevelopmental Disorders, 5*(1), 1–7. https://doi.org/10.1186/1866-1955-5-11

Sucksmith, E., Roth, I., & Hoekstra, R. A. (2011). Autistic traits below the clinical threshold: Re-examining the broader autism phenotype in the 21st century. *Neuropsychology Review, 21*(4), 360–389. https://doi.org/10.1007/s11065-011-9183-9

Sumby W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America, 26*(2), 212–215. https://doi.org/10.1121/1.1907309

Tick, Bolton, P., Happé, F., Rutter, M., & Rijsdijk, F. (2016). Heritability of autism spectrum disorders: a meta-analysis of twin studies. Journal of Child Psychology and Psychiatry, 57(5), 585–595. https://doi.org/10.1111/jcpp.12499

Tye-Murray N., Sommers, M., & Spehar, B. (2007). Auditory and visual lexical neighborhoods in audiovisual speech perception. *Trends in Amplification, 11*(4), 233–241. https://doi.org/10.1177/1084713807307409

Virkud Y., Todd, R. D., Abbacchi, A. M., Zhang, Y., & Constantino, J. N. (2009). Familial aggregation of quantitative autistic traits in multiplex versus simplex autism. *American Journal of Medical Genetics Part B, Neuropsychiatric Genetics, 150B*(3), 328–334. https://doi.org/10.1002/ajmg.b.30810

Warren R. (1970). Perceptual Restoration of Missing Speech Sounds. *Science (American Association for the Advancement of Science), 167*(3917), 392–393. https://doi.org/10.1126/science.167.3917.392