Washington University in St. Louis
Washington University Open Scholarship

Arts & Sciences Electronic Theses and Dissertations

Arts & Sciences

Fall 12-10-2019

Investigating the Relationship Between Gaze Behavior and Audiovisual Benefit Across Various Speech-to-Noise Ratios

Lauren Gaunt
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

 Part of the Cognitive Psychology Commons

Recommended Citation

Gaunt, Lauren, "Investigating the Relationship Between Gaze Behavior and Audiovisual Benefit Across Various Speech-to-Noise Ratios" (2019). *Arts & Sciences Electronic Theses and Dissertations*. 1975.
https://openscholarship.wustl.edu/art_sci_etds/1975

This Thesis is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS
Department of Psychological & Brain Sciences

Investigating the Relationship Between Gaze Behavior and Audiovisual Benefit Across Various Speech-to-Noise Ratios

by

Lauren Taylor Gaunt

A thesis presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Master of Arts

December 2019
St. Louis, Missouri

© 2019, Lauren Taylor Gaunt

St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to sincerely thank my advisor, Mitchell Sommers, for introducing me to audiovisual speech research and for his encouragement of my interests. I would also like to thank my other committee members, Jonathan Peelle and Kristin Van Engen, for their time and investment in my success.

I would like express my appreciation for the helpful feedback, assistance with data collection, and support provided by the members of the Speech and Hearing Lab. I would especially like to thank my research assistants, Madison Kraemer, Layna Paraboschi, and Cameron Tate, for their hard work and dedication to this project.

Thank you to my dear friends, Taylor Levine, Steven Dessenberger, Drew McLaughlin, and Kathryn Hilger for their endless support, understanding, and encouragement. I am particularly grateful to Joe Surman for his encouragement and confidence in me every step of the way.

Last, but certainly not least, I must express my profound gratitude to my parents and my numerous family members for their unfailing support, love, and faith in me. Without them, this accomplishment would not have been possible.

Lauren Gaunt

Washington University in St. Louis

December 2019

Dedicated to my parents, whose love is continually expressed in all modalities.

ABSTRACT OF THE THESIS

Investigating the Relationship Between Gaze Behavior and Audiovisual Benefit Across Various Speech-to-Noise Ratios

by

Lauren Taylor Gaunt

Master of Arts in Psychological & Brain Sciences

Washington University in St. Louis, 2019

Professor Mitchell Sommers

Speech perception improves when listeners are able to see as well as hear a talker, compared to listening alone. This phenomenon is commonly referred to as audiovisual (AV) benefit (Sommers et al., 2005). According to the Principle of Inverse Effectiveness (PoIE), the benefit of multimodal (e.g. audiovisual) input should increase as unimodal (e.g. auditory-only) stimulus clarity decreases. However, recent findings contradict the PoIE, indicating that it should be reassessed. One method for investigating the factors that contribute to AV speech benefit is to examine listeners' gaze behavior with eye tracking. The present study compared young adults' (*N*=50) gaze behavior during AV speech presentations across a range of signal-to-noise ratios in order to determine the relationship between speech-to-noise ratio, gaze behavior, and audiovisual benefit. Participants completed the Build-A-Sentence (BAS) Test, a closed-set test in which participants are asked to identify 3 target words in sentences. Stimuli were presented in auditory-only and audiovisual conditions across four speech-to-noise ratios. Findings were considered from the perspective of the PoIE, which predicts that participants' AV benefit will increase as the auditory signal becomes less intelligible. Additionally, participants' rank order of AV benefit relative to other participants' was compared across speech-to-noise ratios in order to examine individual differences. Participants' AV benefit was consistent with the PoIE, such that AV benefit increased as auditory-only intelligibility decreased. Additionally, participants increased the amount of time spent fixating on the talker's mouth as speech-to-noise ratio decreased.

However, gaze behavior was not a significant predictor of audiovisual benefit, and differences between participants' AV benefit were inconsistent across speech-to-noise ratios. These findings have important implications for research on factors contributing to AV benefit and individual differences in AV benefit.

Chapter 1: Introduction

In adverse listening conditions, speech perception is improved when listeners can both see and hear a talker in comparison to listening alone (Arnold & Hill, 2001; Erber, 1975; MacLeod & Summerfield, 1987; Middelweerd & Plomp, 1987; Sommers, Tye-Murray & Spehar, 2005; Sumby & Pollack, 1954; Van Engen, Xie, & Chandrasekaran, 2017). This phenomenon is commonly referred to as the audiovisual (AV) speech advantage, or *AV benefit* (Sommers et al., 2005). Sumby and Pollack (1954), for example, examined AV benefit by comparing participants' ability to identify test words in auditory-only (AO) and audiovisual (AV) conditions across a range of speech-to-noise ratios (SNRs). Unsurprisingly, speech intelligibility decreased in the AO condition as listening conditions became more challenging. However, speech intelligibility improved considerably when participants were also able to see the talker in the AV condition. Additionally, the difference between AO and AV performance increased as listening conditions became more challenging. Specifically, as auditory speech became less intelligible, listeners increasingly benefited from being able to see as well as hear a talker (Sumby & Pollack, 1954).

1.1 The Principle of Inverse Effectiveness and individual differences in audiovisual benefit

The observation that AV benefit is greater under more difficult listening conditions has been used as supporting evidence for the Principle of Inverse Effectiveness (PoIE) (Stein & Meredith, 1993). The PoIE states that the benefit of multi-modal (AV) compared with unimodal (auditory-only or visual-only) presentations increases as unimodal perception becomes more difficult. Electrophysiological examples of this relationship can be found in single- and population-level

neural responses to multisensory stimuli. Specifically, spike counts and event-related potentials in areas of the brain associated with AV speech perception are greater following multi-modal presentations, compared to unimodal presentations (Meredith & Stein, 1983). This larger response, interpreted as the benefit of multi-modal input, has also been found to have an inverse relationship with auditory and visual stimulus clarity (Meredith & Stein, 1983; Stein et al., 2009; Stevenson et al., 2012). Therefore, the increasing difference between neural responses to multi-modal versus unimodal stimuli with decreasing stimulus clarity is seen as an appreciating benefit of multi-modal input as unimodal signal clarity decreases (Meredith & Stein, 1983). This relationship between multi-modal benefit and stimulus difficulty in neural responses is associated with the inverse relationship between behavioral measures of audiovisual benefit and unimodal stimulus clarity (Stevenson et al., 2012). Together, findings from electrophysiological and behavioral research suggest that the PoIE correctly predicts an increasing benefit of AV presentations as unimodal encoding becomes more difficult.

Although some findings support the PoIE, results from other studies provide contradictory evidence (Ross et al., 2007; Tye-Murray et al., 2010; Stevenson et al., 2015). Tye-Murray et al. (2010) measured younger and older adult participants' ability to correctly identify words in sentences during unimodal and AV presentations. Test stimuli consisted of recordings of a talker saying sentences from the Build-A-Sentence (BAS) Test (Tye-Murray et al., 2008), as well as the CUNY test (Boothroyd et al. 1985). The BAS Test is a closed-set test in which participants are asked to identify target words in sentences. Target words are selected from a list of 36 potential target words and inserted into two possible sentence structures (e.g. "The *team* watched the *moose* and the *girl*", "The *boy* and the *wolf* watched the *cow*"). Participants have access to the list of 36 potential target words throughout the test, and are prompted to respond on

each trial by repeating the sentences aloud. The CUNY Test is an open-set test consisting of lists of unrelated sentences, with each list containing about 100 target words. Participants respond by repeating each sentence out loud, and must repeat words verbatim in order for the response to be scored as correct. To manipulate auditory speech intelligibility, the researchers selected speech-to-noise ratios (SNRs) for each participant to create easy and hard auditory conditions. Videos were presented either unfiltered or with 98% of the visual contrast removed to create easy and hard visual conditions. Participants completed auditory-only, visual-only, and audiovisual trials using all combinations of the unimodal stimuli. Participants' auditory enhancement was compared across all conditions. Auditory enhancement is a calculation of the difference between participants' percent correct scores in Audiovisual and Auditory-Only testing conditions, while also accounting for the percent improvement available (AV – AO) / (1 – AO). The results were compared with what would be expected given the PoIE. Findings consistent with the PoIE would indicate that auditory enhancement increases as either auditory or visual speech perception becomes more difficult. Instead, auditory enhancement decreased in conditions in which either auditory clarity, visual clarity, or both were reduced, compared to conditions with favorable auditory SNRs and visual clarity. These findings challenge previous findings that showed an inverse relationship between AV benefit and unimodal stimulus clarity (Meredith & Stein, 1983; Stein et al., 2009; Stevenson et al., 2012). Furthermore, the PoIE would predict that older adults, whose auditory-only (AO) and visual-only (VO) recognition skills are worse than those of younger adults, would benefit more from AV speech than would younger adults (Pederson et al., 1991; Dancer et al., 1994; Sommers et al., 2005). However, the opposite was true; in conditions with reduced visual contrast, older adults showed less of an AV speech advantage than younger adults (Tye-Murray et al., 2010). In contrast, younger adults and older adults showed similar

benefit for trials with unfiltered video. The observation that participants in both age groups received more AV benefit when the auditory and visual signals were more favorable, combined with age differences when visual contrast was reduced, suggests that the PoIE should be reassessed.

Furthermore, the age differences observed when visual contrast was reduced have implications for individual differences in AV benefit. While AV benefit may vary between individuals (Grant, 2002), the results of Tye-Murray et al. (2010) suggest that detection of individual differences in AV benefit may depend on testing conditions. Tye-Murray et al. (2010) found that age differences in AV benefit were only detected when the visual signal was less clear; this likely means that our ability to detect individual differences in AV benefit is largely dependent on the conditions in which people are tested.

1.2 Individual differences in YA audiovisual benefit and gaze behavior

Whereas AV benefit is a well-established phenomenon, there are differences in the degree to which individuals benefit from AV speech presentations compared to unimodal presentations. Although this presented as age differences in Tye-Murray et al. (2010), differences in AV benefit are found even within a homogenous sample of younger adults with normal hearing (MacLeod & Summerfield, 1990). MacLeod and Summerfield (1990) asked twenty participants to identify words in sentences using AO and AV conditions with background noise and measured the minimum SNR needed for participants to correctly identify at least 3 words in each sentence. The difference between minimum SNRs for AO and AV conditions was interpreted as the amount that participants benefited from the addition of visual speech information. Although all

participants benefited from AV speech compared to AO speech, this gain ranged widely, with some participants gaining as little as 2.7 dB and others gaining as much as 9.5 dB.

If there are individual differences in AV benefit, it is natural to ask what factors contribute to people's ability to benefit from AV speech. One approach to identifying these factors is use of eye tracking to examine participants' gaze behavior during AV speech presentations in a variety of listening conditions. This methodology addresses the question of which areas of the talker's face participants focus on during AV speech perception, and does this gaze behavior change when the auditory signal is more difficult to identify? Buchan et al. (2008) sought to answer these questions by collecting eye tracking data while participants viewed videos of talkers saying low-context sentences (e.g., "Mrs. White would consider the mold") from the Speech in Noise (or SPIN) sentences (Kalikow, Stevens & Elliott, 1977). Participants were presented with AV stimuli in a Noise Absent condition and a Noise Present condition. In the Noise Present condition, noise was added to degrade the auditory signal and reduce participant performance to 40.0% correct in the Noise Present condition, compared to 96.8% in the Noise Absent condition. Participants' gaze behavior in the Noise Absent and Noise Present conditions was compared. The results indicated that in the Noise Present condition, participants spent more time fixating on the talker's mouth and nose and reduced the amount of time fixating on the talker's eyes compared to in the Noise Absent condition. Unfortunately, the experiment did not include an auditory-only (AO) condition and therefore it is unclear whether the observed gaze behavior would have helped to produce an AV benefit. However, the differences in gaze behavior between the Noise Present and Noise Absent conditions suggest that listeners can adjust their gaze behavior as a strategy for overcoming a noisy speech signal (Buchan et al., 2008).

Buchan et al. (2008) emphasized that while a great deal of visual speech information

comes from a talker's mouth, listeners may also attend to areas of a talker's face that supply social cues, such as the eyes. The need to monitor social cues on a talker's face may explain why listeners focused more on the talker's eyes in the Noise Absent condition compared to the Noise Present condition. Without noise, a listener can rely more heavily on the auditory signal for speech information and use the visual signal to monitor social cues as well as speech cues; however, with noise present, the listener must increase time spent focusing on the mouth in order to compensate for a degraded auditory signal. If this is the case, then listeners may vary the amount of time spent fixating on a talker's mouth as the need to compensate for a noisy speech signal varies.

Attempts to use eye tracking to investigate individual differences in gaze behavior during audiovisual speech perception include studies examining individual differences in susceptibility to the McGurk illusion (Gurler et al., 2015). In the McGurk illusion, an auditory syllable is dubbed onto a different visual syllable (e.g., an auditory /bɑ/ paired with a visual /gɑ/). For some, integration of these incongruent auditory and visual cues affects their overall speech perception, causing them to perceive an illusory syllable (such as /dɑ/, in the previous example). Perception of this illusory stimulus is typically referred to as the McGurk effect (McGurk & MacDonald, 1976). However, there are individual differences in susceptibility to the McGurk effect, with some participants being far less likely to perceive this illusion than are others (McGurk & MacDonald, 1976; Nath & Beauchamp, 2012; Brown et al., 2018). People who are more susceptible to the McGurk effect have been thought to be more adept at combining, or *integrating*, auditory and visual speech cues, whereas those who were less susceptible were thought to be less skilled at integrating these cues. However, more recent research demonstrates that McGurk susceptibility is not correlated with audiovisual benefit, indicating that

measurement of susceptibility to the McGurk illusion does not equate to measurement of the ability to integrate congruent AV speech (Van Engen et al., 2017; Hickok et al., 2018). For this reason, McGurk susceptibility and audiovisual benefit should be viewed as measuring different aspects of integrating auditory and visual inputs, rather than as informing the same integration process.

Although concerns about the lack of correlation between McGurk susceptibility and AV benefit are valid, there are compelling similarities between findings from studies that have used eye tracking to examine individual differences in either McGurk susceptibility or AV speech advantage. For example, Gurler et al. (2015) used eye tracking while participants were presented with McGurk stimuli in an identification task and found a relationship between individual differences in the amount of time spent focusing on the talker's mouth and McGurk susceptibility. Specifically, participants who reported fewer McGurk-like percepts spent less time focusing on the talker's mouth. The researchers suggested that this relationship between gaze behavior and McGurk susceptibility was found because participants who spent less time focusing on the talker's mouth were more likely to miss important visual speech cues (Gurler et al., 2015).

Recent findings also indicate that individual differences in gaze behavior are related to individual differences in AV benefit (Alsius et al., 2016). Alsius et al. (2016), for instance, used eye tracking to compare the gaze behavior of individuals who benefit most (high gain) and least (low gain) from the addition of visual speech cues with a range of visual clarity. Participants viewed videos of a talker in AV conditions with varied spatial frequency (i.e. blur) of the talker's image. Participants were sorted into high gain and low gain groups based on their AV benefit in the condition in which the talker's image was not blurred. The researchers hypothesized that high

gain and low gain participants differed in their ability to benefit from AV speech because of differences in gaze behavior. In this case, differences in gaze behavior were interpreted as a sign of difference in skill at extracting visual speech cues. The results confirmed that there were group differences in gaze behavior. High gain participants spent more time fixating on the talker's mouth when presented with words than did low gain individuals. This suggests that differences in gaze behavior may be a contributing factor to individual differences in AV speech benefit, such that individuals who spend more time fixating on a talker's mouth receive more of an AV speech advantage than do participants who spend less time fixating on the mouth (Alsius et al., 2016). Given Buchan and colleagues' (2008) argument that participants' shift in gaze behavior across listening conditions is a strategy to compensate for a noisy signal, it would seem that high gain participants are better at using this strategy than are low gain individuals.

Alsius et al. (2016) had also predicted that high gain participants would be more dependent on visual speech cues and that their AV speech advantage would decrease as the visual clarity of the image decreased. Indeed, there appeared to be a relationship between participants' AV gain in the unfiltered condition and their gain as visual clarity decreased, such that high gain participants received less of an audiovisual benefit than did low gain participants as visual clarity decreased (Alsius et al., 2016). The researchers suggested that AV benefit in conditions with decreased visual clarity, compared to a condition with an unfiltered image, can provide an index of the extent to which individuals rely on visual speech information.

1.3 The Present Study

Observations of gaze behavior during presentations of AV speech stimuli demonstrated that people's gaze behavior shifts when noise is added to the auditory signal (Buchan et al., 2005), and that high-gain and low-gain individuals exhibit different gaze behavior during AV speech

presentations (Alsius et al., 2016). The present study aims to compare the gaze behavior of high gain and low gain participants across a range of SNRs. This will be addressed by using eye tracking to compare gaze behavior during the BAS test in AO and AV conditions, with stimuli presented in quiet and three different SNRs. Three SNRs were used in this task because although gaze behavior has been investigated during AV speech perception, previous studies have only done so using one auditory SNR (Buchan et al., 2008; Alsius et al., 2016). Using a range of SNRs allows us determine whether participants increasingly focus on a talker's mouth as auditory speech becomes less intelligible, as well as examine the relationship between gaze behavior and individual differences in AV benefit. Using listening conditions of varied difficulty also allows a test of the PoIE, which predicts that participants' AV benefit will increase as SNR decreases.

Listeners' gaze behavior was compared across a range of SNRs to determine whether SNR affected gaze behavior, and if gaze behavior predicted individual variability in AV benefit. We predicted that as auditory noise increased, participants would increase the amount of time spent fixating on the talker's mouth and that increased fixation time on the mouth would correspond with increased AV benefit.

Chapter 2: Method

2.1 Participants

Young adult participants (*N*=50, Females = 35, Mean age = 19.1 years, range = 18-21 years, *SD* = 0.9 years) were recruited from Washington University's Psychology Subjects Pool. All participants were native English speakers with no known hearing disorders and normal or corrected-to-normal vision. All participants were screened to have at least 20/40 visual acuity using the Snellen eye chart. Participants who reported use of corrective vision were asked to wear it during the vision screening and during the experiment. Participants provided informed consent and received course credit as compensation for participation in accordance with the Washington University Institutional Review Board.

2.2 Test Stimuli

Test stimuli included sentences from the Build-a-Sentence (BAS) test, a sentence recognition test that uses a set of target words inserted into a consistent sentence structure (Tye-Murray et al, 2008). In this test, target words for each sentence are selected without replacement from a closed set of 36 nouns and placed in two possible sentence frames (for example, "The *boys* and the *dog* watched the *mouse*.", or "The *snail* watched the *girls* and the *whale*."). All nouns in the list refer to animate objects with eyes. Following each trial, participants were presented with the list of possible target words on the computer screen and prompted to respond by repeating the appropriate target words aloud. The BAS test was ideal for the current study because it is designed to avoid the ceiling effects that can occur with AV sentence tests and also prevents participants from relying on context in order to identify target words. Each BAS list consisted of 12 sentences, which were randomized so that each BAS list included all 36 words once. Digital

recordings of 16 BAS lists were used for the current study. Eight of the generated lists were used for the four audiovisual (AV) conditions, and eight were used for the four auditory-only (AO) conditions. Scoring was based on the number of correctly identified target words in each trial, regardless of the order in which participants repeated them.

Stimulus recordings were prepared from digital video recordings of the face and neck of a female, native-North American English speaker reading the lists of sentences as they appeared on a teleprompter. The stimuli were leveled using Adobe Audition to ensure that the auditory portion of all stimuli had about the same RMS amplitude. For experimental conditions, speech-shaped noise was added to the auditory stimulus using MATLAB to reduce the intelligibility of the speech signal. Auditory stimuli were created for three speech-in-noise conditions with different speech-to-noise ratios (SNR's). Speech-in-noise stimuli were created with the goal of producing about 30%, 50%, and 70% correct identification of target words in the AO condition.

Pilot testing was used to determine that SNR's of -12, -9, and -6 dB were appropriate to achieve the desired response accuracy. The speech signal was generated at -23.9 dB, and noise was generated at -11.9, -14.9, and -17.9 dB to create stimuli for -12, -9, and -6 SNR conditions, respectively.

2.3 Procedure

Participants completed a demographic questionnaire and vision screening assessment prior to completing the experimental task. Afterward, participants were seated in a sound-proofed booth facing an EyeLink 1000 eye tracker and a computer monitor. Equipment setup was done in accordance with Eyelink specifications. The experimenter first explained the procedure to the participant, and then performed a 13-point calibration and validation. In order for a participant's eye tracking data to be included for the analysis, the maximum allowed average error was 1.0 visual degrees, and the maximum error on a single fixation point was 1.5 visual degrees during

validation. Once calibration was complete, participants were asked to minimize movement for the duration of the experiment. Participants completed practice trials prior to beginning data collection to ensure that they understood the procedure and knew how to appropriately respond on each trial.

During the experimental task, the auditory signal was played through headphones at a comfortable listening level. On AV trials, the video signal was displayed on the computer monitor. On AO trials, a fixation cross was displayed in the middle of the screen during stimulus presentation. After stimulus presentation, a response screen appeared with a list of all 36 possible target words. Participants were instructed to provide a spoken response only for each trial when this screen appeared. An audio recorder was used to record participants' verbal responses. Testing was self-paced; after responding on each trial, participants progressed to the next trial by hitting the "Space" bar on a keyboard. All stimuli were presented using PsychoPy.

Stimuli were presented in 8 randomized blocks (4 AO and 4 AV), with each block consisting of stimuli from 2 BAS lists at a single SNR (a total of 24 trials per block). Trials within each block were presented in a random order. Eye tracking data were collected for all AV trials using the Eyelink 1000 eye-tracking system. In total, the procedure took about 50 minutes.

2.4 Analysis

2.4.1 Scoring

Participants' responses were noted by native English speakers from audio recordings of each participant's verbal responses. Responses were then compared to the correct stimulus target words and scored using a script in R Studio. Participants received 1 point for each correctly identified target word, with a maximum of 3 points possible for each trial. Noun pluralization was ignored when marking responses as correct or incorrect.

2.4.2 Audiovisual benefit

To calculate participants' audiovisual benefit, we compared the percentage of words correctly identified in each of the AO and AV conditions. Participants were given an AO score and an AV score for each SNR condition, with each score representing the percentage of words correctly identified in the respective block. Because we were specifically interested in the benefit of adding a visual speech signal to an auditory speech signal, AV benefit was calculated by subtracting participants' AO scores from their AV scores (AV-AO).

2.4.3 Gaze behavior

To analyze gaze behavior during audiovisual trials, we compared the amount of time spent fixating on four areas of interest (AOI's) during audiovisual trials. Four rectangular AOI's were created (Mouth, Nose, Right Eye, Left Eye) using SR Research Data Viewer (Figure 1). To allow comparison of fixations on the talker's eyes and mouth, the total area of the right and left eye AOI's was equal to the area of the mouth AOI. The mouth AOI was created so that it included all parts of the talker's mouth when it was open at its widest point. A fifth, elliptical AOI for the talker's entire face was also created, in order to determine how much participants looked at the talker's face.

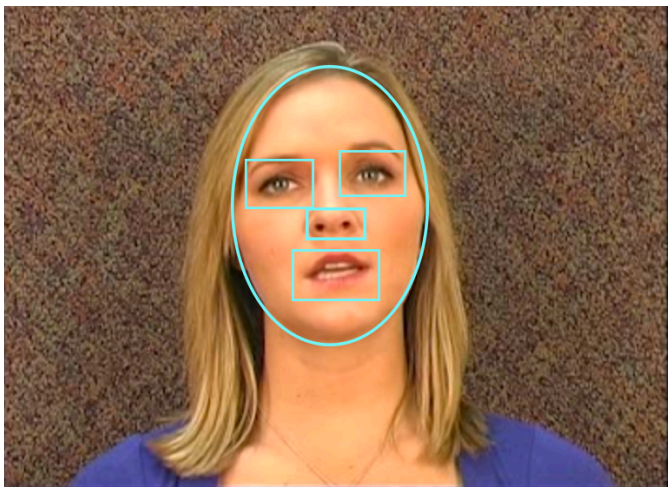

Figure 2.1 Areas of interest (AOIs) were created around the talker's mouth, nose, eyes, and face for analysis.

Reports for individual participants' gaze behavior (including number of fixations and total dwell time for each AOI) were acquired using SR Research Data Viewer, and analysis was conducted using R Studio. To compare gaze behavior during AV presentations of stimuli with varying SNR's, the average fixation time in milliseconds for each AOI during every block of AV presentations was calculated by participant.

2.4.4 Modeling

Data were analyzed using mixed effects hierarchical regression models using the lme4 package in R v.1.2.1335. All models included participant as a random effect, to account for individual differences in audiovisual benefit. In models using average fixation time as the dependent variable, SNR was included as a fixed effect. Average fixation time, SNR, and the interaction between average fixation time and SNR were included as fixed effects in models predicting audiovisual benefit. Separate analyses were used for each AOI to analyze the effects of average fixation time on AV benefit. Pairwise comparisons of models were conducted using likelihood ratio tests and Bayes Factors using the BIC.

Chapter 3: Results

3.1 Auditory-only and audiovisual performance

Figure 3.1 shows the mean percent words correct in auditory-only (AO) and audiovisual (AV) conditions for each SNR. The addition of noise for each SNR condition was effective in lowering the average percent correct in auditory-only conditions. On average, participants correctly identified 95.6 % of words correctly in the no-noise condition; 78.9% in the -6 SNR condition; 58.8% in the -9 SNR condition, and 42.4% in the -12 SNR condition. When they received AV presentations, participants on average correctly identified 96.6% of words correctly in the no-noise condition; 86.1% in the -6 SNR condition; 75.6% in the -9 SNR condition, and 65.1% in the -12 SNR condition.

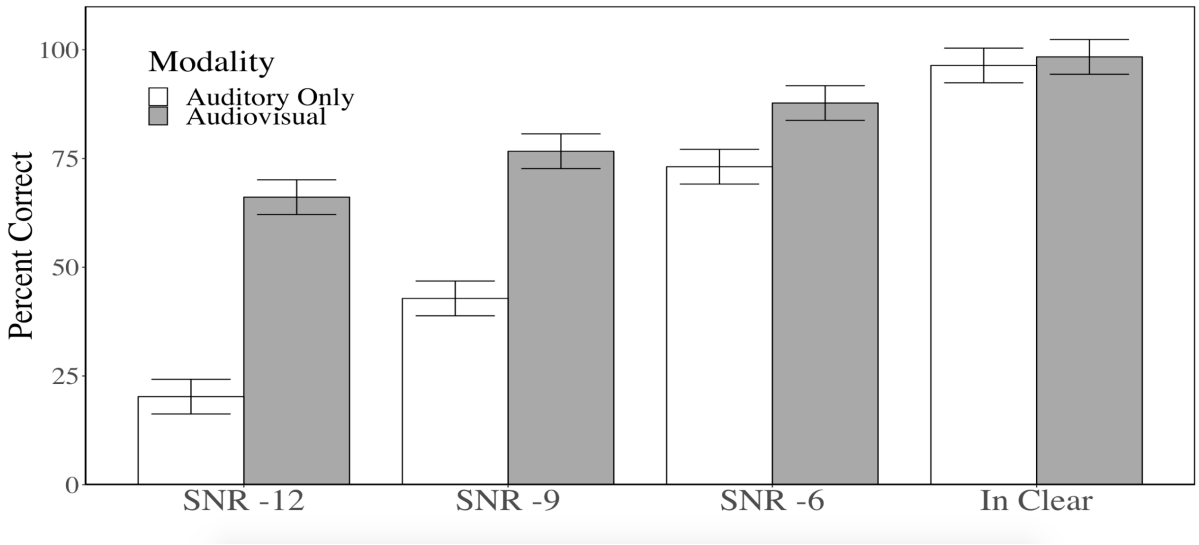

Figure 3.1. Mean percent correct responses in Auditory-Only and Audiovisual conditions per speech-to-noise ratio (SNR). Error bars indicate the 95% confidence interval.

3.2 Audiovisual benefit

Figure 3.2 shows the mean AV benefit (AV-AO) for each SNR condition. We first tested the

unconditional model, with participant included as a random affect. We then tested for the fixed effect of SNR against the baseline random effect of participant on AV benefit. SNR was a significant predictor in the model (χ^2= 148.91; $p < 0.001$), with AV benefit increasing as a function of SNR. Specifically, participants' AV benefit increased from a mean of 17.80% (SD=4.18%) in the -6 SNR condition to 34.56% (SD=4.05%) in the -9 SNR condition and was highest in the -12 SNR condition (*M*=45.86%, SD=4.01%).

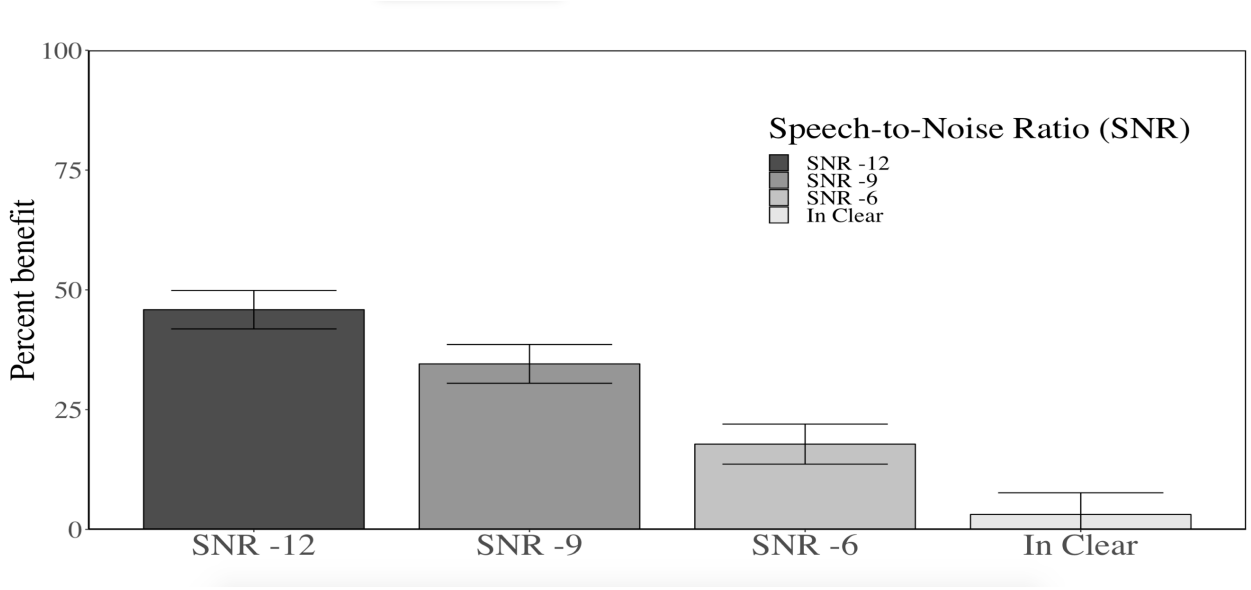

Figure 3.2. Mean audiovisual benefit (AV-AO) per speech-to-noise ratio (SNR). Error bars indicate the 95% confidence interval.

3.3 Gaze behavior

Figure 3.3 shows the descriptive statistics for the percentage of fixation time for all areas of interest (AOIs) for each speech-to-noise ratio (SNR). To test the relationship between percentage of fixation time, AOI, and SNR, stepwise comparisons of mixed effects hierarchical regression models were used. We first tested an unconditional model predicting percentage of fixation time

and included participant as a random effect. We then added the AOI grouping variable as a fixed effect. AOI was a significant predictor of average percentage fixation time (χ^2=6878.1; p < 0.001). As shown by Figure 3.3, participants spent the most time fixating on the talker's mouth (M= 35.11%, SD=28.63%), followed by the talker's nose (M=13.16%, SD=18.29%), left eye (M=6.42%, SD=13.23%) and right eye (M=1.76%, SD=6.63%). We then included SNR as an additional fixed effect. Based on a lower BIC score and a low Bayes Factor, adding SNR as a fixed effect did not improve the model (χ^2= 22.311; p < 0.001, BIC= -9339.0, BF_{10}= 0.7485). We then added the interaction of AOI and SNR as a fixed effect. The model comparison was significant (χ^2= 610.47; p < 0.001), with strong evidence in favor of the model including the interaction (BF_{10}=5.334*10^{112}). Table 3.1 shows regression coefficients, standard error, and p-values for the interaction model. This model also supported our hypothesis that average fixation time would vary by AOI, and that this relationship would change across SNRs.

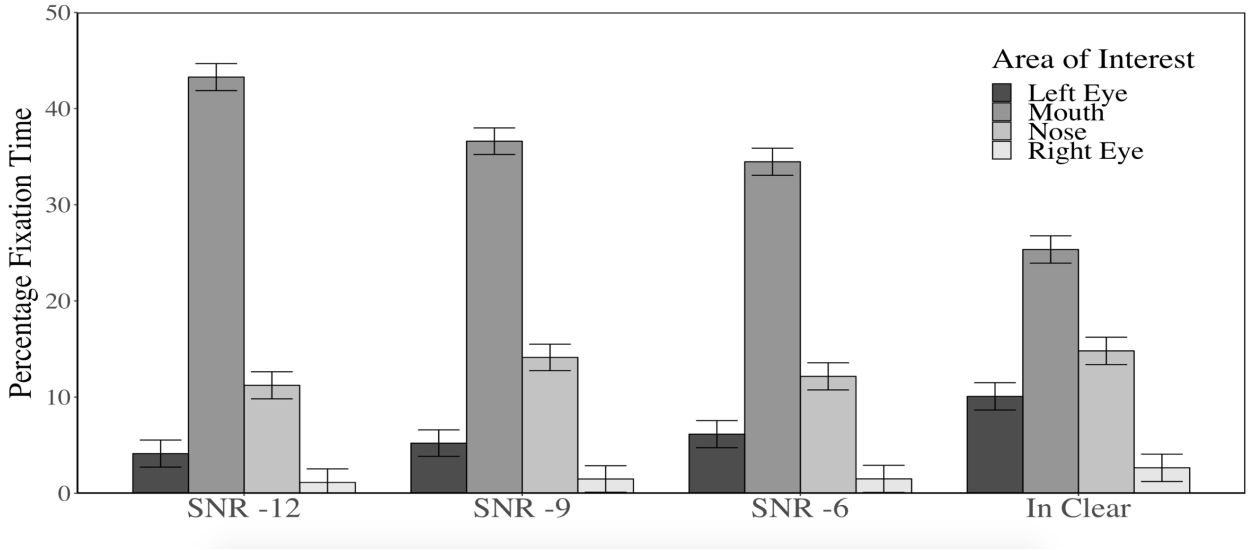

Figure 3.3. Mean percent of time spent fixating on each Area of Interest (AOI) per speech-to-noise ratio (SNR). Error bars indicate the 95% confidence interval.

Table 3.1 Regression coefficients, standard error, and p-values for the Area of Interest (AOI) and Speech-to-Noise Ratio (SNR) interaction model predicting percentage fixation time.

	Estimate (ß)	*t*-value	*p*
Fixed Effects			
Right Eye	2.64	3.65	< 0.001
Mouth	25.34	34.96	< 0.001
Nose	14.80	20.42	< 0.001
Left Eye	10.07	13.90	< 0.001
-6 SNR	-1.14	-1.45	0.15
-9 SNR	-1.16	-1.49	0.14
-12 SNR	-1.51	-1.92	0.055
Mouth * -6 SNR	10.27	9.18	< 0.001
Nose * -6 SNR	-1.50	-1.34	0.18
Left Eye * -6 SNR	-2.78	-2.49	0.013
Mouth * -9 SNR	12.42	11.29	< 0.001
Nose * -9 SNR	0.48	0.44	0.66
Left Eye * -9 SNR	-3.69	-3.36	< 0.001
Mouth * -12 SNR	19.44	17.39	< 0.001
Nose * -12 SNR	-2.05	-1.84	0.07
Left Eye * -12 SNR	-4.43	-3.96	<0.001

Linear contrast comparisons of the interaction model were conducted post-hoc using the Tukey method. As shown in Table 3.2, the average fixation time was significantly different for all AOI pairs, regardless of SNR. Participants consistently spent more time fixating on the talker's mouth than any other AOI, and increased the percentage of time spent fixating on the talker's mouth as noise increased. Percentage of fixation time on both of the talker's eyes decreased as noise increased, and the percentage of time spent fixating on the talker's nose was variable across SNRs.

Table 3.2 Linear contrast comparisons for Areas of Interest (AOI) and Speech-to-Noise Ratio (SNR) interaction model.

		Estimate	SE	*p*
In Quiet				
	Right eye - Mouth	-22.70	0.797	<.001
	Right eye - Nose	-12.16	0.797	<.001
	Right eye – Left eye	-7.43	0.797	<.001
	Mouth – Nose	10.54	0.797	<.001
	Mouth – Left eye	15.27	0.797	<.001
	Nose – Right eye	4.73	0.797	<.001
-6 SNR				
	Right eye - Mouth	-32.97	0.785	<.001
	Right eye - Nose	-10.66	0.785	<.001
	Right eye – Left eye	-4.65	0.785	<.001
	Mouth – Nose	22.32	0.785	<.001
	Mouth – Left eye	28.32	0.785	<.001
	Nose – Right eye	6.01	0.785	<.001
-9 SNR				
	Right eye - Mouth	-35.12	0.758	<.001
	Right eye - Nose	-12.64	0.758	<.001
	Right eye – Left eye	-3.73	0.758	<.001
	Mouth – Nose	22.48	0.758	<.001
	Mouth – Left eye	31.39	0.758	<.001
	Nose – Right eye	8.91	0.758	<.001
-12 SNR				
	Right eye - Mouth	-42.14	0.784	<.001
	Right eye - Nose	-10.10	0.784	<.001
	Right eye – Left eye	-3.00	0.784	<.001
	Mouth – Nose	32.05	0.784	<.001
	Mouth – Left eye	39.15	0.784	<.001
	Nose – Right eye	7.10	0.784	<.001

3.4 Gaze behavior and audiovisual benefit

Different models were used to test the relationship between the fixed effects of average fixation time for each AOI with SNR and audiovisual benefit.

3.4.1 Mouth AOI

To test the relationship between time spent fixating on the talker's mouth, SNR, and audiovisual benefit, we used stepwise comparisons of mixed effects hierarchical regression models. First, we used an unconditional model with participant as a random effect and AV benefit as the dependent variable. We then added average mouth fixation time as a fixed effect. Mouth fixation time was a significant predictor of AV benefit, with increased time spent fixating on the talker's mouth predicting increased AV benefit (χ^2= 5.7; $p < 0.017$). We then tested a model with SNR added as a fixed effect. SNR was a significant predictor of AV benefit, and improved the model (χ^2= 152.58; $p < 0.001$, BF_{10}=3.4637*10^{30}). After controlling for SNR, mouth fixation time was not a significant predictor of AV benefit.

We then included the interaction between mouth fixation time and SNR as a fixed effect. Table 3.3 shows regression coefficients, standard error, and *p*-values for the interaction model. The interaction was not a significant predictor (χ^2= 4.48; p =.21). This, along with a small Bayes Factor (BF_{10}=0.1453007), suggested that adding the interaction between SNR and mouth fixation time did not significantly improve the model. Therefore, our hypothesis that AV benefit would vary as a function of both SNR and time spent fixating on the talker's mouth was not supported. The relationship between average mouth fixation time and AV benefit for each SNR is shown in Figure 3.4.

Table 3.3 Regression coefficients, standard error, and *p*-values for the Mouth Fixation Time and Speech-to-Noise ratio (SNR) interaction model predicting percentage fixation time.

	Estimate (ß)	*t*-value	*p*
Fixed Effects			
Intercept	1.69	1.07	0.28
Mouth fixation time	$-0.71*10^{3}$	-0.62	0.53
-6 SNR	7.74	3.35	< 0.00
-9 SNR	0.13	5.47	< 0.00
-12 SNR	0.23	9.41	< 0.00
Mouth fixation time * -6 SNR	$-0.74*10^{3}$	-0.51	0.61
Mouth fixation time * -9 SNR	$0.17*10^{2}$	1.23	0.22
Mouth fixation time * -12 SNR	$-0.16*10^{03}$	-1.84	0.07

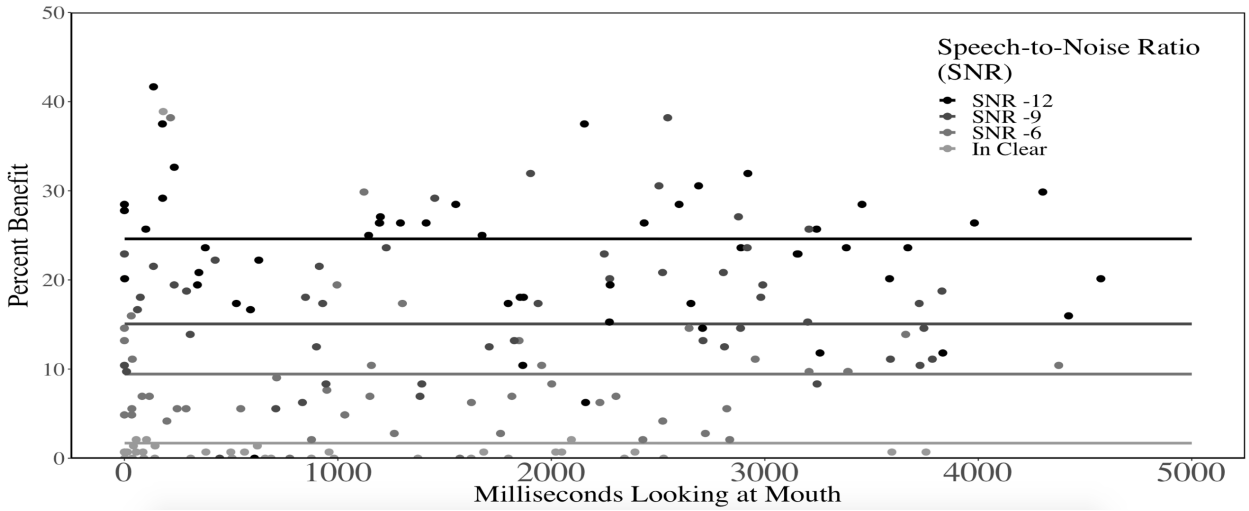

Figure 3.4 Relationship between mouth fixation time and mean percent audiovisual benefit (AV-AO) for each speech-to-noise ratio (SNR).

As an exploratory analysis, we selected the model with only mouth fixation time and SNR as fixed effects as the most appropriate model. This model was used for post-hoc linear contrast comparisons using the Tukey method. As shown in Table 3.4, AV benefit was

significantly different across all levels of SNR, with benefit increasing as noise increased.

Table 3.4 Linear contrast comparisons for model with mouth fixation time and SNR as fixed effects.

	Estimate	SE	*p*
In Clear – -6 SNR	-16.30	1.57	<.001
In Clear – -9 SNR	-22.38	1.59	<.001
In Clear – -12 SNR	-6.55	1.54	<.001
-12 SNR – -6 SNR	-9.75	1.53	<.001
-12 SNR – -9 SNR	-15.83	1.54	<.001
-6 SNR – -9 SNR	-6.08	1.53	<.001

3.4.2 Nose AOI

To test the relationship between nose fixation time, SNR, and AV benefit, we first used an unconditional model with participant as a random effect and AV benefit as the dependent variable. We then added average nose fixation time as a fixed effect. Nose fixation time was not a significant predictor of AV benefit, and did not improve the model (χ^2= 0.115; p =.735, BF_{10} = 2.5697*10^{-32}). We then added SNR to the model as a fixed effect. SNR was a significant predictor of AV benefit, and significantly improved the model (χ^2= 157.91; p <.001, BF_{10} =4.2002*10^{30}).

We then included the interaction between nose fixation time and SNR as a fixed effect. Table 3.5 shows regression coefficients, standard error, and *p*-values for the interaction model. The interaction was not a significant predictor (χ^2= 5.156; p =.1608). This suggested that adding the interaction between SNR and nose fixation time did not significantly improve the model (BF_{10} =0.17155). Because of the results of all model comparisons, the model including only nose fixation time and SNR was selected as the model which best fit the data. The relationship between average nose fixation time and AV benefit for each SNR is shown in Figure 3.5.

Linear contrast comparisons were conducted post-hoc using the Tukey in the model with nose fixation time and SNR as fixed effects. As shown in Table 3.6, all levels of SNR were significantly different from each other, with AV benefit increasing as noise increased.

Table 3.5 Regression coefficients, standard error, and p-values for the Nose Fixation Time and speech-to-noise ratio (SNR) interaction model predicting percentage fixation time.

	Estimate (ß)	*t*-value	*p*
Fixed Effects			
Intercept	1.63	1.12	0.27
Nose fixation time	-0.001	-0.68	0.50
-6 SNR	6.79	3.52	< 0.00
-9 SNR	14.39	7.31	< 0.00
-12 SNR	20.62	10.99	< 0.00
Nose fixation time * -6 SNR	-0.001	-0.51	0.58
Nose fixation time * -9 SNR	0.004	1.22	0.22
Nose fixation time * -12 SNR	0.004	1.29	0.20

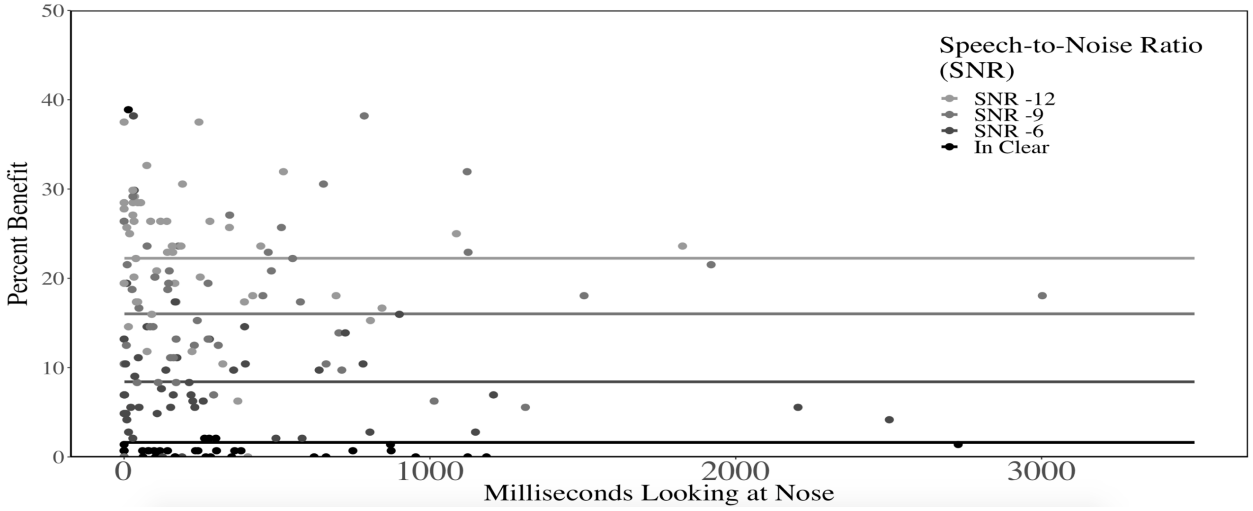

Figure 3.5 Relationship between nose fixation time and mean percent audiovisual benefit (AV-AO) for each speech-to-noise ratio (SNR).

Table 3.6. Linear contrast comparisons for model with nose fixation time and SNR as fixed effects.

	Estimate	SE	*p*
In Clear – -6 SNR	-6.35	1.53	<.001
In Clear – -9 SNR	-15.93	1.52	<.001
In Clear – -12 SNR	-21.97	1.53	<.001
-6 SNR – -9 SNR	-9.58	1.53	<.001
-6 SNR – -12 SNR	-15.62	1.52	<.001
-9 SNR – -12 SNR	-6.03	1.53	<.001

3.4.3 Eyes AOI

The average fixation times for each eye were combined for each block by participant. The combined average fixation time (which we will refer to as "eyes fixation time") was used to test whether eyes fixation time, SNR, or the interaction between eyes fixation time and SNR was a significant predictor of AV benefit. First, we used an unconditional model with participant as a random effect and AV benefit as the dependent variable. We then added eyes fixation time as a fixed effect. Eyes fixation time was a significant predictor of AV benefit (χ^2= 15.163; p <.001) but had a small Bayes Factor (BF_{10} =1.7334*10^{-28}). We then added SNR as a fixed effect, which was a significant predictor of AV benefit and improved the model (χ^2= 143.93; p <.001, BF_{10} =7.4948*10^{28}).

We also tested a model that included the interaction between eyes fixation time and SNR as a fixed effect. Table 3.7 shows regression coefficients, standard error, and *p*-values for the interaction model. Inclusion of the interaction term worsened model fit (χ^2= 11.078; p =.011, BF_{10} =1.3727). For this reason, we selected the model including only eyes fixation time and SNR as fixed effects as the model which best fit the data. The relationship between average eyes

fixation time and AV benefit for each SNR is shown in Figure 3.6.

Linear contrast comparisons were conducted post-hoc using the Tukey in the model with eyes fixation time and SNR as fixed effects. All levels of SNR were significantly different from each other, with AV benefit increasing as noise increased (Table 3.8).

Table 3.7 Regression coefficients, standard error, and p-values for the eyes fixation time and speech-to-noise ratio (SNR) interaction model predicting percentage fixation time.

	Estimate (ß)	*t*-value	*p*
Fixed Effects			
Intercept	1.63	1.12	0.27
Eyes fixation time	-0.001	-0.68	0.50
-6 SNR	6.79	3.52	< 0.00
-9 SNR	14.39	7.31	< 0.00
-12 SNR	20.62	10.99	< 0.00
Eyes fixation time * -6 SNR	-0.001	-0.51	0.58
Eyes fixation time * -9 SNR	0.004	1.22	0.22
Eyes fixation time * -12 SNR	0.004	1.29	0.20

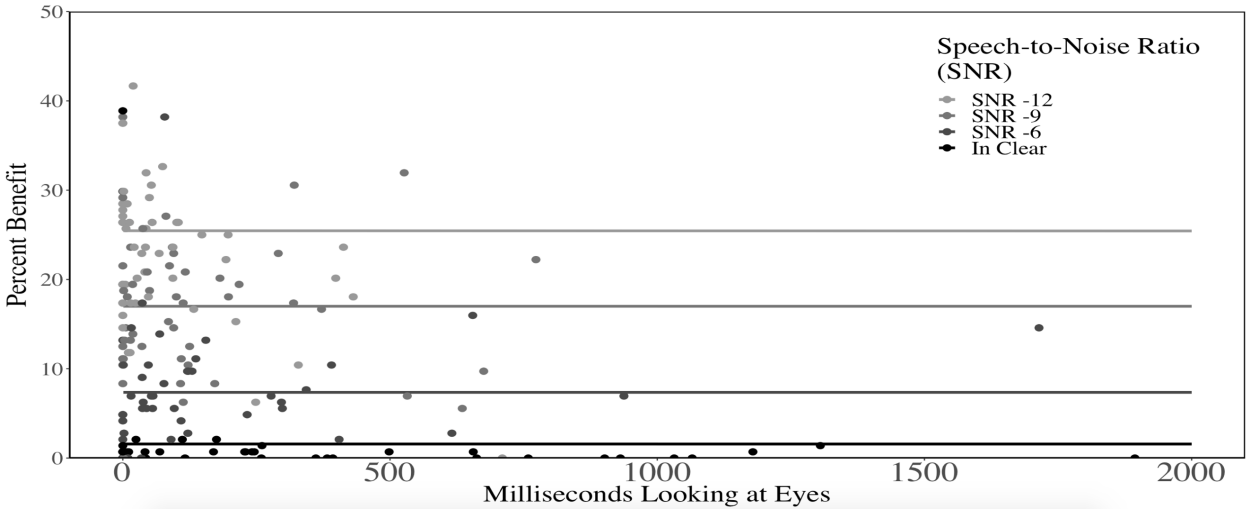

Figure 3.6 Relationship between eyes fixation time and mean percent audiovisual benefit (AV-AO) for each speech-to-noise ratio (SNR).

Table 3.8 Linear contrast comparisons for model with eyes fixation time and SNR as fixed effects.

	Estimate	SE	*p*
In Clear – -6 SNR	-5.95	1.55	0.001
In Clear – -9 SNR	-15.42	1.58	<.001
In Clear – -12 SNR	-21.31	1.61	<.001
-6 SNR – -9 SNR	-9.46	1.53	<.001
-6 SNR – -12 SNR	-15.36	1.54	<.001
-9 SNR – -12 SNR	-5.89	1.53	0.001

3.5 Individual differences in AV benefit

Spearman's rank-order correlations were conducted to determine whether a participant's AV benefit relative to all other participants' benefit correlated across SNRs. The resulting correlation matrix is shown in Table 3.9. Notably, participants' ranks were not highly correlated across SNRs, suggesting that the difference between individual participants' AV benefit was not consistent across a range of listening conditions. Because consistent individual differences in AV benefit were not detected, it was not possible to test our hypothesis that gaze behavior would differ between High Benefit and Low Benefit participants.

Table 3.9 Spearman's rank-order correlations for participant audiovisual benefit by speech-to-noise ratio (SNR).

	-6 SNR	-9 SNR	-12 SNR
-6 SNR	1.00	-	-
-9 SNR	0.22	1.00	-
-12 SNR	0.05	0.23	1.00

Chapter 4: Discussion

4.1 Gaze behavior shifts as auditory noise increases

In line with our hypothesis and previous results (Buchan et al., 2008; Alsius et al., 2016), participants spent more time fixating on the talker's mouth when noise was added to the auditory speech signal. Moreover, there was a systematic increase in time spent focusing on the mouth as SNR became less favorable. Previous studies have examined gaze behavior in only a Noise Absent and Noise Present condition (Buchan et al., 2008), or with consistent auditory noise and a range of visual clarity (Alsius et al., 2016). Our method specifically allowed us to examine participants' gaze behavior across a range of listening conditions. Our results indicated that participants spent more time fixating on the talker's mouth than on the eyes or nose in all listening conditions. Furthermore, as noise increased in the auditory signal, participants increased the amount of time spent fixating on the talker's mouth. Additionally, the percentage of time spent fixating on the talker's eyes decreased as auditory noise increased. These results resemble Buchan and colleagues' (2008) finding that participants increased fixation duration on the talker's mouth and decreased fixation duration on the eyes in a Noise Present condition compared to a Noise Absent condition. Our finding that this change in gaze behavior occurred as a function of SNR supports Buchan and colleagues' (2008) suggestion that gaze behavior reflects prioritization of social cues in the absence of noise and of speech cues when noise is present. When auditory speech is easily intelligible, the auditory signal alone is sufficient to understand what is being said. Therefore, gaze behavior in easy listening conditions may reflect a strategy

whereby the listener can attend to visual social cues from the talker's face, as well as speech cues. When the auditory signal is degraded by noise, however, increased time spent focusing on the talker's mouth may indicate that the listener has shifted their gaze in an attempt to compensate for a noisy signal by attending to visual speech cues. Interestingly, our results indicate that this strategy does not help to improve the AV benefit.

4.2 Audiovisual benefit consistent with Principle of Inverse Effectiveness

The results of the present study supported our hypothesis that participants' AV benefit would be consistent with the PoIE. Specifically, participants increasingly benefitted from the addition of visual speech cues as the auditory speech signal became more degraded. When noise was used to decrease participants' AO scores to about 79% correct in the -6 SNR condition, participants experienced a gain of about 7% in the audiovisual condition. This gain increased to about 17% in the -9 SNR condition, and was highest in the -12 SNR condition, with participants' AV benefit increased to an average of about 27%. Notably, audiovisual benefit is not consistently evaluated with the same approach in the literature. In multiple studies (Sommers et al., 2005; Tye-Murray et al., 2010; Alsius et al., 2016), AV benefit has been calculated by comparing the relative gain of audiovisual speech cues while controlling for baseline unimodal performance [Multimodal score – Unimodal score / 1 – Unimodal score]. This manner of calculating a normalized AV benefit is useful for between-subjects designs because it takes into account the amount of benefit possible in an AV condition compared to a unimodal condition. For example, a participant with an increase from 40% in an auditory-only condition to 75% in an AV condition would have a raw AV benefit of 35%, but a normalized benefit of 58%. However, a participant who increases

from 95% in an auditory-only condition to 98% in an AV condition would also have a normalized benefit of 58%, despite a raw benefit of only 3%. Despite the large difference between the two participants' raw AV benefit, their normalized scores reflect that relative to their auditory-only performance, both participants improved to the same degree in an AV condition.

Although calculating participants' normalized AV benefit scores is useful for comparing participants' gains based on each individual's room for improvement compared to an auditory-only condition, it is challenging to compare the degree to which participants benefit from audiovisual speech input across a range of signal degradation based on normalized scores. For example, a participant whose raw percent benefit scores were 5%, 15%, and 25% in auditory-only conditions with -6 SNR, -9 SNR, and -12 SNR might have a larger normalized AV benefit score in the -6 SNR condition, despite having a much larger raw gain in the -12 SNR condition. In this case, the participant's raw AV benefit scores are consistent with the PoIE, but their normalized benefit scores are contradictory to it. This is a potential explanation for why some studies' results are contradictory to the PoIE, while the present study's findings support it. In the present study, we were particularly interested in differences in AV benefit across a range of auditory signal degradation. For this reason, we chose to calculate participants' raw AV benefit scores in order to make their benefit across SNRs more comparable and found an inverse relationship between AV benefit and auditory signal clarity, supporting the PoIE

.4.3 Gaze behavior and audiovisual benefit

We hypothesized that participants' AV benefit and amount of time spent fixating on the talker's mouth would increase as noise in the auditory signal increased. These hypotheses were supported, as both mouth fixation time and AV benefit did increase as a function of SNR.

However, our results contradict our hypothesis that the interaction between mouth fixation time and SNR would be related to AV benefit. After controlling for SNR, the interaction was not a significant predictor of AV benefit. In light of the coincident increase in AV benefit and mouth fixation time as noise increased, this finding is surprising.

In the literature, one explanation offered to explain how participants benefit from audiovisual compared to auditory-only speech cues is that visual speech contains articulatory cues that offer complementary information when the auditory speech signal is degraded (Grant, Walden, & Seitz, 1998; Summerfield, 1987). If participants truly benefit because of these complementary cues, it would be expected that participants who spend the most time fixating on the talker's mouth, and therefore have the most opportunity to take advantage of visual articulatory cues, would show more AV benefit. However, we found that mouth fixation time was not predictive of AV benefit after controlling for SNR, suggesting that participants' AV benefit was unlikely to have arisen from complementarity. Despite participants' tendency to increase time spent fixating on the talker's mouth in noisy listening conditions, this change in gaze behavior was not a main contributor to their AV benefit. This supports an alternative explanation that listeners' benefit results from temporal information provided by the visual speech stimulus, rather than articulatory information. Such temporal information can serve as an attentional cue and indicate to a listener when in time a talker is speaking, and when the auditory signal should provide speech information.

4.4 Individual differences in audiovisual benefit

Although a main goal of the present study was to investigate the relationship between gaze behavior and individual differences in AV benefit, we did not detect consistent individual differences in AV benefit. Participants' AV benefit relative to that of other participants' was

inconsistent across SNRs (Table 9), and we were unable to classify individuals as being consistently either high benefit or low benefit. Because individual differences in AV benefit were not detected in this sample, we were unable to compare gaze behavior between high benefit and low benefit individuals. However, based on our finding that mouth fixation time was not a significant predictor of AV benefit and the dispersion of our data for mouth fixation time, we suspect that gaze behavior would not differ significantly between high benefit and low benefit individuals.

Our inability to detect individual differences in AV benefit in the present study may be due to a limitation of our method. Multiple studies (Tye-Murray et al., 2010; Alsius et al., 2016) have detected individual or age differences in AV benefit when visual clarity, rather than auditory clarity, was manipulated. It is possible that individual differences could have been detected in the current study if visual clarity had also been manipulated within subjects.

4.5 Conclusion

The finding that the amount of time spent fixating on a talker's mouth is not a good predictor of audiovisual benefit has important implications for research on audiovisual benefit. Despite the emphasis in past research on participants' fixations on a talker's mouth, the results of the present study suggest that this focus will not provide useful information regarding how participants benefit from audiovisual speech compared to auditory-only speech input. Our results also emphasize the importance of testing conditions in attempts to identify individual differences in AV speech benefit. Specifically, our finding that participants' ranked AV benefit was inconsistent across SNRs differs from the finding in Alsius et al. (2016) that High Benefit participants consistently benefited more than Low Benefit participants across a range of visual degradation. This indicates that consistent individual differences in AV benefit may be more

detectable when visual clarity is manipulated than when auditory clarity is varied. Replication of the present study's results, as well as the addition of elements such as working memory measures, a visual-only condition, and manipulated visual clarity may provide useful insight into the sources and consistency of individual variability in the ability to benefit from added visual speech input. Furthermore, it is possible that individual differences were inconsistent in the present study due to the homogeneity of the sample. It is possible that individuals with age-related hearing loss, for whom the auditory speech signal is less reliable, are more reliant on visual speech input and may show differences in AV benefit and gaze behavior compared to a sample of healthy younger adults. A study comparing the AV benefit and gaze behavior of older adults and younger adults may find age differences in gaze behavior, as well as its relationship with AV benefit.

References

Alsius, A., Wayne, R. V., Paré, M., & Munhall, K. G. (2016). High visual resolution matters in audiovisual speech perception, but only for some. *Attention, Perception, & Psychophysics*, *78*(5), 1472–1487. doi: 10.3758/s13414-016-1109-4

Arnold P., & Hill F. (2001). Bisensory augmentation: a speechreading advantage when speech is clearly audible and intact. *Br. J. Psychol. 92,* 339–355. 10.1348/000712601162220

Boothroyd, A., Hanin, L., & Hnath-Chisholm, T. The CUNY Sentence Test. New York: City University of New York; 1985.

Brown VA, Hedayati M, Zanger A, Mayn S, Ray L, Dillman-Hasso N, et al. (2018) What accounts for individual differences in susceptibility to the McGurk effect? *PLoS ONE* 13(11): e0207160. https://doi.org/10.1371/journal.pone.0207160

Buchan, J. N., Paré, M., & Munhall, K. G. (2008). The effect of varying talker identity and listening conditions on gaze behavior during audiovisual speech perception. *Brain Research*, *1242*, 162–171. doi: 10.1016/j.brainres.2008.06.083

Dancer, J., Krain, M., Thompson, C., Davis, P., & Glenn, J. (1994). A cross-sectional investigation of speechreading in adults: Effects of age, gender, practice and education. *Volta Review*, *96*, 31-40.

Erber, N. P. (1975). Auditory-visual perception of speech. *J. Speech Hear. Disord. 40,* 481–492. 10.1044/jshd.4004.481

Grant, K. W. (2002). Measures of auditory-visual integration for speech understanding: A theoretical perspective. *Journal of the Acoustical Society of America*, *112*, 30–33.

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-

visual integration. *Journal of the Acoustical Society of America*, *103,* 2677–2690.

Gurler, D., Doyle, N., Walker, E., Magnotti, J., & Beauchamp, M. (2015). A link between individual differences in multisensory speech perception and eye movements. *Attention, Perception, & Psychophysics*, *77*(4), 1333–1341. doi: 10.3758/s13414-014-0821-1

Hickok, G., Rogalsky, C., Matchin, W., Basilakos, A., Cai, J., Pillay, S., … Fridriksson, J. (2018). Neural networks supporting audiovisual integration for speech: A large-scale lesion study. *Cortex*, *103*, 360–371. doi: 10.1016/j.cortex.2018.03.030

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*, *61*(5), 1337–1351. doi: 10.1121/1.381436

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. doi: 10.1038/264746a0

MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *Br. J. Audiol., 21,* 131–141. 10.3109/03005368709077786

Macleod, A., & Summerfield, Q. (1990). A procedure for measuring auditory and audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use. *British Journal of Audiology*, *24*(1), 29–43. doi: 10.3109/03005369009077840

Meredith, M., & Stein, B. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, *221*(4608), 389–391. doi: 10.1126/science.6867718

Middelweerd, M. J., & Plomp, R. (1987). The effect of speechreading on the speech-reception threshold of sentences in noise. *The Journal of the Acoustical Society of America*, *82*(6), 2145–2147. doi: 10.1121/1.395659

Nath AR, Beauchamp MS. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *Neuroimage*. 2012; 59: 781–787. pmid:21787869

Pederson KE, Rosenthal U, Moller MB. Longitudinal study of changes in speech perception between 70 and 81 years of age. *Audiology*. 1991; 30:201–211.

Ross L. A., Saint-Amour D., Leavitt V. M., Javitt D. C., & Foxe J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environment. *Cereb. Cortex, 17,* 1147–1153. 10.1093/cercor/bhl024

Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear Hear., 26,* 263–275. 10.1097/00003446-200506000-00003

Stein, B. E., & Meredith, M. A. (1993). *The Merging of the Senses.* Cambridge, MA: MIT Press.

Stein, B. E., Stanford, T. R., Ramachandran, R., Perrault, T. J., & Rowland, B. A. (2009). Challenges in quantifying multisensory integration: alternative criteria, models, and inverse effectiveness. *Experimental Brain Research*, *198*(2-3), 113–126. doi: 10.1007/s00221-009-1880-8

Stevenson, R. A., Bushmakin, M., Kim, S., Wallace, M. T., Puce, A., & James, T. W. (2012). Inverse effectiveness and multisensory interactions in visual event-related potentials with audiovisual speech. *Brain Topography*, *25*(3), 308–326. doi: 10.1007/s10548-012-0220-7

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am*., *26,* 212–215. 10.1121/1.1907309

Stevenson, R. A., Nelms, C. E., Baum, S. H., Zurkovsky, L., Barense, M. D., Newhouse, P. A., & Wallace, M. T. (2015). Deficits in audiovisual speech perception in normal aging

emerge at the level of whole-word recognition. *Neurobiology of Aging*, *36*(1), 283–291. doi: 10.1016/j.neurobiolaging.2014.08.003

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lipreading*, 3–51. Hillsdale: Erlbaum.

Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., Hale, S., & Rose, N. S. (2008). Auditory-visual discourse comprehension by older and young adults in favorable and unfavorable conditions. *International Journal of Audiology*, *47*(sup2). doi: 10.1080/14992020802301662

Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., & Hale, S. (2010). Aging, audiovisual integration, and the principle of inverse effectiveness. *Ear and Hearing*, 1. doi: 10.1097/aud.0b013e3181ddf7ff

Van Engen, K. J., Xie, Z., & Chandrasekaran, B. (2017). Audiovisual sentence recognition not predicted by susceptibility to the McGurk effect. *Attention, Perception, & Psychophysics*, *79*(2), 396–403. doi: 10.3758/s13414-016-1238-9