Washington University School of Medicine
Digital Commons@Becker

Independent Studies and Capstones | Program in Audiology and Communication Sciences

2012

Lipreading difficulty during audiovisual integration

Stephanie Shapiro Wieczorek
Washington University School of Medicine in St. Louis

Follow this and additional works at: https://digitalcommons.wustl.edu/pacs_capstones

 Part of the Medicine and Health Sciences Commons

Recommended Citation

Wieczorek, Stephanie Shapiro, "Lipreading difficulty during audiovisual integration" (2012). *Independent Studies and Capstones.* Paper 657. Program in Audiology and Communication Sciences, Washington University School of Medicine.
https://digitalcommons.wustl.edu/pacs_capstones/657

This Thesis is brought to you for free and open access by the Program in Audiology and Communication Sciences at Digital Commons@Becker. It has been accepted for inclusion in Independent Studies and Capstones by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.

LIPREADING DIFFICULTY DURING AUDIOVISUAL INTEGRATION

by

Stephanie Shapiro Wieczorek

A Capstone Project
submitted in partial fulfillment of the
requirements for the degree of:

Doctor of Audiology

Washington University School of Medicine
Program in Audiology and Communication Sciences

May 17, 2013

Approved by:
Brent Spehar, Ph.D., Capstone Project Advisor
Nancy Ty-Murray, Ph.D., Second Reader

Abstract: Audiovisual integration ability for word-level stimuli was assessed using two talkers, one easy to lipread and the other hard to lipread. No significant effect for integration ability was found for the two talkers.

copyright by

Stephanie Shapiro Wieczorek

2013

Acknowledgements

I would like to extend my sincerest appreciation to Brent Spehar, Ph.D. for sharing his time, knowledge, and expertise with me throughout the completion of my Capstone project. Without his guidance and extensive work in programming and data analysis, this investigation would not have been possible. I have gained invaluable experience while working with him. I would also like to thank Nancy Tye-Murray, Ph.D. for her help with the editing and revision of this manuscript and for giving me the opportunity to learn more about research through work in her lab. I genuinely appreciate all of the hard work and the dedication to the students that I have seen from the faculty and staff of the Program in Audiology and Communication Sciences. Special thanks go to all those who participated in this study.

I would also like to thank my classmates for the moral support and friendship they have shown me during my time at Washington University. Finally, special thanks go to my husband and my family for the endless amount of love, support, and help that they have given me throughout my life.

Table of Contents

List of Tables and Figures

Abbreviations

AE	Auditory Enhancement
ASHA	American Speech-Language-Hearing Association
AV	Audiovisual
A-only	Auditory Only
BAS	Build-A-Sentence
CUNY	City University of New York
CV	Consonant-Vowel
dB	Decibel
HL	Hearing Level
Hz	Hertz
IE	Integration Enhancement
PC	Personal Computer
POIE	Principle of Inverse Effectiveness
SD	Standard Deviation
SNR	Signal-to-Noise Ratio
TDH	Telephonics Dynamic Headphone
T1	Talker 1
T2	Talker 2
VCV	Vowel-Consonant-Vowel
VE	Visual Enhancement
V-only	Visual Only

Introduction

Technological advancements in recent years have significantly improved the ability of hearing aids to enhance speech sounds and manage background noise. While many hearing aid users have found improvement in their ability to understand speech, many do not find enough benefit from hearing aids alone. Aural rehabilitation can play a major role in helping these hearing aid users improve their listening ability and increase their communication effectiveness. Aural rehabilitation combines the use of assistive listening devices, knowledge and understanding of hearing loss, and speechreading, along with strategies for environmental modification, listening, and breakdown strategies to improve the quality of communication for hearing aid users. A main communication strategy used to improve speech understanding is the utilization of the visual speech cues available when watching the talker's face. The use of speechreading cues and lipreading can potentially provide large gains in speech perception with the added benefit that it costs the user little or nothing.

It is well known that speech perception can be affected by observing a speaker's face. The "McGurk Effect" presents an excellent example of how visual cues influence the ability to identify what is being said (McGurk & MacDonald, 1976). In their famous study it was found that under certain circumstances if one stop consonant was presented auditorily and a second consonant differing in only place was presented at the same time visually, a third consonant, often medial to the other two, would be perceived. For example, "da" is often perceived when the combination of the auditory "ba" and visual "ga" are presented simultaneously. The McGurk Effect is powerful evidence for how listeners merge two streams of speech information into a single percept, even if they are providing conflicting information.

As opposed to the artificial situation created during the McGurk Effect, real life situations

involve the merging of auditory and visual signals that are complimentary. When this is the case, and the auditory signal is poor, looking at the speaker's face can significantly enhance speech perception ability. This is especially important for those with hearing loss and when there is distortion of the auditory signal by background noise, hearing impairment, or poor listening environments. Sumby and Pollack (1954) described how the addition of visual speech to a severely degraded auditory signal could improve performance by as much as 80 percentage points. Similarly, the perception of speech features has been proven to be more precise when both auditory and visual speech cues are available relative to auditory cues alone (Grant, Walden, & Seitz, 1998). Grant, Tufts, and Greenberg (2007) showed that place of articulation is the most accessible feature from visual speech. Green and Kuhl (1991) studied reaction times for audiovisual speech segments that differed across a continuum of place and voicing. They found that the addition of a perceivable second feature (i.e. voicing + place) varied reaction times relative to a single feature alone. They showed that people receive greater benefit when both the auditory (voicing) and visual (place) signals are present. Findings were consistent with others that have also shown that the audiovisual benefit is likely due to the complementary nature of the information coming from the two modalities (Grant & Seitz, 1998; Grant et al., 1998; Summerfield, 1987).

The benefit associated with combining auditory and visual cues has been well established. It is, however, important to distinguish between what is the *result* of the combination of the information coming from the two modalities and the *process* of combining the two modalities. Enhancement can be thought of as the *result* and is described as the benefit received from the addition of a second modality, typically to either auditory or visual stimuli (Sumby & Pollack, 1954). Integration is the *process* that is used to combine information across modalities (Grant,

2002). For example, measures of audiovisual integration typically assess how much additional benefit is seen in auditory-visual speech recognition above what is expected or predicted based on speech recognition ability for each of the individual unimodal inputs (Tye-Murray et al., 2010). Although enhancement measures have often been used to describe the auditory-visual speech benefit, attempts to use these measures as indices of audiovisual integration are often confounded under uncontrolled circumstances. For example, when measuring the enhancement associated with adding visual speech to auditory speech it is not possible to be sure what amount of benefit is due to lipreading ability and what amount is due to integration ability.

As mentioned above, much is known about the potential benefit that comes from combining auditory and visual speech information. Surprisingly little is known, however, about the nature of the integration mechanism itself. For example, there have been conflicting reports regarding how the information available in the stimuli will affect the ability to integrate the auditory and visual stimuli. At least three potential models of integration have been described.

Grant & Seitz (1998) have proposed that the audiovisual (AV) integration ability of an individual is independent of the ability to extract auditory and visual cues. This implies that individuals have an innate ability to integrate these cues, no matter the how difficult it is to see and hear them. Grant & Seitz used low-context sentences, vowel-consonant-vowel (VCV) segments, and consonant-vowel (CV) segments, all presented in noise, to measure AV integration in forty-one listeners. In order to measure the effect of discrepant auditory and visual information on the ability to integrate the stimuli, they compared conditions using congruent and incongruent audio and visual stimuli. Results showed that significant benefit from visual speech cues was demonstrated for consonant and sentence recognition in noise. However, there were considerable differences in the amount of AV integration obtained by participants, with no

relationship found between the amount of AV integration for sentences stimuli and that of the consonant stimuli.

In contrast to a model that would suggest integration ability is independent of the amount of auditory and visual input, other research on multi-sensory integration by Lakatos, Chen, O'Connell, Mills, and Schroeder (2007) describes a principle of inverse effectiveness (PoIE). The PoIE states that the ability to integrate two unimodal inputs (including auditory speech and visual speech) should be enhanced when one of the inputs is relatively small or degraded. In other words, as information from one modality becomes sparser, the amount of information gleaned from the two inputs should be more than expected. To measure this principle, macaque monkeys were presented with auditory clicks of differing intensities (20-80 dB) and somatosensory stimuli (mild electrical stimulation of the median nerve) that remained at a constant level. Both types of stimuli were presented individually and simultaneously to determine neural responses for all conditions. Results indicated that multisensory enhancement was greatest in conditions where auditory click stimuli were presented at a lower intensity level. These findings support the hypothesis that the stimuli level (difficulty) affects integration ability.

Also in contrast to a model that suggests integration is a static ability, recent research by Tye-Murray et al. (2010) compared the amount of audiovisual integration for speech measured at different levels for the unimodal inputs. In that study, a closed-set word recognition task in sentence format (Build-A-Sentence Test) and an open-set sentence recognition task (CUNY Sentence Test) were used in the presence of background noise and with two levels of visual contrast in the video signal to assess integration abilities in a total of 106 normal hearing adults at varying levels of A-only and V-only input. Results indicated that integration was highest when the unimodal inputs from the auditory and visual channels were easier to perceive as

compared to the conditions in which the inputs from the auditory and visual channels were relatively harder to perceive. These findings were in contrast to those predicted by the both the PoIE and Grant & Seitz (1998).

Taken together, results from previous studies are inconclusive regarding the nature and amount of audiovisual integration that occurs when the levels of the unimodal inputs vary. The current study attempted to address this issue by measuring integration ability using varying amounts of input from the visual modality while holding auditory performance constant. This approach built upon the research conducted by Tye-Murray et al (2010). In that study, a single talker was used to assess all levels of the unimodal conditions and the combined audiovisual conditions. The levels of input from the visual modality were manipulated using high-contrast and low-contrast signals while the two levels of auditory input were varied by controlling two levels of background noise. As mentioned above, using the manipulation of video contrast showed that integration was partially dependent on unimodal input levels. The current study used a similar method with an easy-to-lipread speaker and a hard-to-lipread speaker.

In summary, the results from recent studies investigating the relationship between audiovisual integration and the varying amounts of unimodal input suggests one of three possible outcomes for the current study. If no difference in integration are shown across the two talkers, results would be in support of the theory proposed by Grant & Seitz (1998) in which the amount of integration is independent of the unimodal inputs. This would indicate that integration is a constant, or even an ability, that can be predicted regardless of the amount of input. Results indicating more integration for the hard-to-lipread talker (Lakatos et al., 2007) would support the POIE, which proposes that those conditions with more difficult unimodal inputs will show increased integration amounts. This would indicate that audiovisual integration is increased

when less information is available. Finally, if results show that more integration occurred with the easy-to-lipread talker, the findings from Tye-Murray et al. (2010) would be replicated. These results would indicate that we are able to integrate better when more information is available.

Methods

Participants

The study protocol was approved by the Human Research Protection Office at Washington University School of Medicine (#201110160). Participants were recruited through the use of fliers and all spoke English as their first language. All participants were at least 18 yeas of age and informed consent was obtained from each individual prior to beginning the study. One testing session was needed, lasting approximately 1.5 hours. Participants were not compensated for their time.

Twenty-four young adults (mean age 23.76 years; range 21-27 years; SD = 1.50; 22 females, 2 males) qualified for and participated in the investigation. Table 1 shows each participant's age and gender. Participants were screened to have corrected or uncorrected 20/40 visual acuity or better and normal visual contrast sensitivity using the Snellen eye chart and the Pelli-Robson contrast sensitivity test, respectively (Pelli et al, 1998). Anyone with a history of central nervous system disorder was disqualified from participating in the study. Hearing acuity was screened by presenting pure-tones at 25 dB HL for frequencies including 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 3000 Hz, 4000 Hz, 6000 Hz, and 8000 Hz. Hearing screenings were completed using a calibrated Madsen Auricle audiometer and TDH-49 headphones. Participants were asked to sit in a sound treated booth and press a handheld button when they heard the tones. Participants were disqualified if they were unable to hear any of the tonal stimuli at 25 dB HL.

Participant	Sex	Age (Years;Months)	Years of Education
2	Male	25;3	17
3	Female	22;2	16.5
4	Male	27;2	18
5	Female	26;9	19
6	Female	23;10	17
7	Female	22;5	16.5
8	Female	23;5	18
9	Female	22;8	16.5
10	Female	24;8	18.5
11	Female	25;1	18.5
12	Female	24;6	18.5
13	Female	22;5	16.5
14	Female	22;6	16.5
15	Female	21;8	15.5
16	Female	22;6	17.5
17	Female	22;4	16.5
18	Female	25;10	19
19	Female	24;1	17.5
20	Female	24;0	17.5
21	Female	23;1	16.5
22	Female	23;8	17
23	Female	22;9	16.5
24	Female	25;0	18.5
25	Female	22;7	16.5

Table 1. Participant demographic information.

Stimuli

A modified version of the Build-a-Sentence (BAS) test was used during this experiment. The BAS test is a closed-set matrix test that assesses word identification for 36 words in sentence context (Tye-Murray et al., 2008). Table 2 lists the words and sentence formats used in the BAS test. The BAS recordings from Tye-Murray et al. (2008) used a single female talker for all presentations. In the modified version used here, recordings of two different female speakers

presenting the BAS stimuli were taken from a study that looked at lipreading across multiple talkers (Tye-Murray, Spehar, Myerson, Hale, & Sommers; Submitted). The two talkers, reported here as Talker 1 and Talker 2, were determined to be "easy–to-lipread" and "hard-to-lipread" respectively. Lipreading scores from the nine young normally hearing participants in that study were at 52.8 % correct when watching Talker 1 and 28.3 % for Talker 2.

Six lists of 12 sentences were created for each talker. The audio and video materials used for the stimuli were high-quality digital recordings of the talkers speaking sentences in a General American English dialect. Each list contains the same 36 words randomly assigned to the sentence structures shown in Table 2. Three lists were used for practice and setting the level of the background noise. The other three were assigned to one of three conditions: auditory-only (A-only), visual-only (V-only), and audiovisual (AV). The video stimuli were edited using Adobe Premiere Elements software to ensure consistency in size and the head-and-shoulders framing of the participant. Adobe Audition software was used to level the audio for each sentence. The results were stimuli equal in loudness across sentences and across the two talkers. A program written in LabView was designed and created specifically for this test.

Please choose one of these types of sentences:

The ___ and the ___ watched the ___ and the ___.
The ___ and the ___ watched the ___.
The ___ watched the ___ and the ___.
The ___ watched the ___.

Please choose all words from this list:

bear	cat	deer	fawn	geese	men	saint	team	whale
bird	cook	dog	fish	girls	mice	seal	toad	wife
boys	cop	dove	fox	goat	mole	snail	tribe	wolf
bug	cow	duck	frog	guest	moose	son	troop	worm

Table 2. Response screen shown to participants between each presentation of the BAS test. Taken from Tye-Murray et al. (2008).

Procedures

After completing the hearing and vision screenings, participants were asked to sit in a sound treated room approximately 0.5 meters from a 17" ELO touch-systems monitor. Stimuli were presented via PC (Dell, Precision) configured for dual-screen presentation. The screen used to present the stimuli was located in the testing booth while the second screen was located outside the booth for experimenter use. Audio portions of the stimuli were routed from the PC audio card to a calibrated Madsen Auricle audiometer. Changes in signal-to-noise ratio (SNR) were controlled via LabView software and a Tucker Davis Technologies real-time processor (RP2). Auditory presentations were made through two speakers orientated at +/- 45 degrees to the left and right of the participant when looking at the monitor. Stimulus-specific calibration noise was used to check the system's calibration prior to each testing session. During stimuli presentations, each talker produced one of the four possible sentence constructions. A-only, V-only, and AV conditions were used. All stimuli were presented in the presence of background noise consisting of four-talker babble. The verb "watched" was used in each sentence with either one or two key words preceding and following the verb. After each sentence trial, participants were shown a screen that displayed all four potential sentence types along with the 36 key word options available to fill-in the blanks. The response screen can be seen in Table 2. Participants were required to respond by repeating aloud the sentence they heard. Before testing began, participants were informed that no one word could be used twice within a single sentence. Guessing with a test-appropriate response was required when participants were uncertain about a sentence.

Testing was completed in three portions. Participants were first given 12 practice items to familiarize them with the task. Practice included four sentences from each condition (A-only,

V-only, and AV) that alternated between each talker. Practice was conducted at a +5 SNR. Participants were encouraged to ask questions during the practice if needed to ensure full understanding of the task.

After completing the practice items a procedure to set the background noise levels to be used during the scored portion of testing was completed. For testing, the noise was held constant at 62 dB SPL and the speech was adjusted to reflect changes in SNR. Before the procedure began, an additional six A-only sentences were presented at -10 and -15 SNRs (two presentation at -10 and one at -15 for each talker) to give participants practice in more difficult listening conditions. To set the SNRs participants were asked to respond to 60 sentences (2.5 lists for each talker) at five SNR levels, ranging from -20 to 0 in increments of 5, in the A-only condition. A modified ASHA SRT (ASHA, 1988) procedure was used to obtain responses at each SNR level. Results for each participant were then used to generate a psychometric function showing percent correct word identification relative to SNR level for each talker. Two SNRs needed to correctly identify 30% of the key words were interpolated based on the psychometric function; one SNR for each talker. The 30% criterion was used in order to avoid ceiling performance in the AV condition while also avoiding floor performance in the A-only condition.

In the final, scored portion of testing, participants were asked to respond to 72 sentences (3 lists for each talker), which included an equal mix of A-only, V-only, and AV stimuli. For each talker, auditory stimuli were set at the pre-determined SNR level. The stimuli conditions were presented in random order, with presentations alternating between talkers. All participants received the same stimuli in the same order.

Results

Raw scores from Participant 24 were more than 3 standard deviations from the means and consequently were not used in the statistical analysis of the data. The data from the remaining 23 participants (mean age 23.71 years, range 21-27 years, SD = 1.51, 21 females, 2 males) were analyzed. Results for scores in the three conditions (A-only, V-only, and AV) and associated correlation coefficients are reported first. Methods of calculation and results for the measures of audiovisual benefit and audiovisual integration are then reported with associated correlation coefficients.

Performance in A-only, V-only, and AV Conditions

Figure 1 shows A-only mean percent correct scores for each talker. On average, in the A-only condition, participants correctly identified 31% of words for Talker 1 and 36% of words for Talker 2. A repeated measures ANOVA, looking at within-subjects differences for the two talkers, indicated no difference in A-only scores between talkers ($F(1, 22) = 1.26$; $p = .273$). The procedure for controlling A-only performance across the two talkers, therefore, appears to have been effective. A-only performance was not correlated for the two talkers ($r = -.068$).

Figure 2 shows V-only mean percent correct scores for each talker. On average, participants were able to correctly identify 46% of the words for Talker 1 and 27% for Talker 2. A repeated measures ANOVA indicated a difference between the two talkers for the lipreading scores ($F(1, 22) = 55.2$; $p < .0001$) with the mean percent correct score for Talker 1 nearly 20 percentage points higher than that of Talker 2. V-only scores were similar to those found for the same talkers in Tye-Murray et al. (Submitted). V-only performance for the two talkers was highly correlated ($r = .698$; $p = .0001$).

Figure 3 shows AV mean percent correct scores for each talker. On average, participants were able to correctly identify 74% and 63% of the words for Talker 1 and Talker 2 respectively. A repeated measures ANOVA indicated a difference between the two talkers for the AV scores ($F(1, 22) = 37.8$; $p < .0001$) with performance higher for Talker 1 than for Talker 2. AV performance for the two talkers was also correlated ($r = .481$; $p = .0191$).

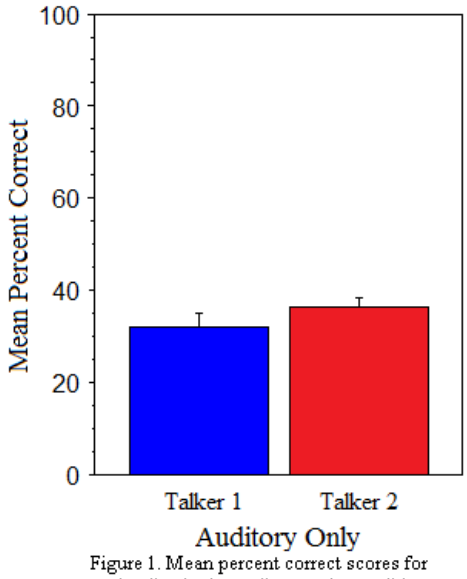

Figure 1. Mean percent correct scores for each talker in the auditory only condition.

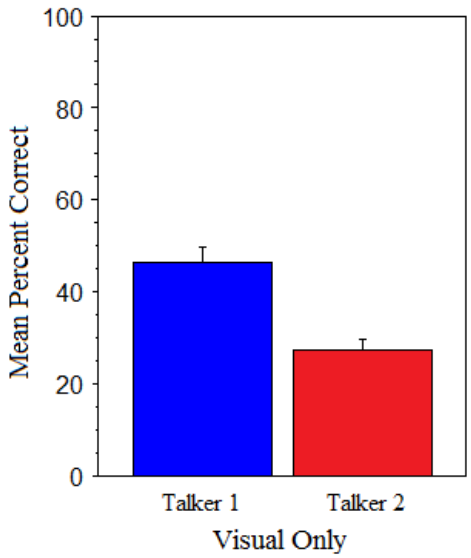

Figure 2. Mean percent correct scores for each talker in the visual only condition.

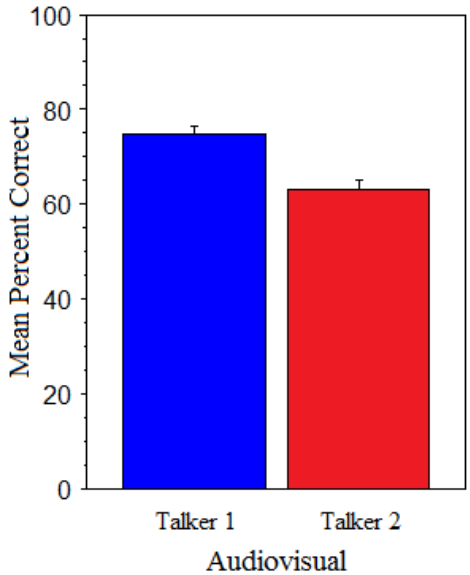

Figure 3. Mean percent correct scores for each talker in the audiovisual condition.

Audiovisual Benefit

Visual enhancement (VE) can be described as the normalized amount of benefit resulting from the addition of a visual speech signal to an auditory speech signal (Sommers et al., 2005), where VE = (AV – A) / (1 – A). The amount of VE is assessed by first subtracting the raw A-only score from the raw AV score, which provides an index of the non-normalized benefit of adding the visual signal to the auditory signal. The difference is then divided by the result of subtracting the raw A-only score from 100 percent. This normalizes for the amount of improvement that adding the visual signal could have potentially provided. The result is typically expressed as the percent of the possible improvement that was achieved when the visual signal is added. Figure 4 presents the average VE scores for each talker. A repeated measures ANOVA indicated a difference between the two talkers for VE ($F(1, 22) = 63.6$; $p < .0001$) with participants showing greater VE for Talker 1 (62%) than for Talker 2 (41%). VE scores were correlated for the two talkers ($r = .529$; $p = .0085$).

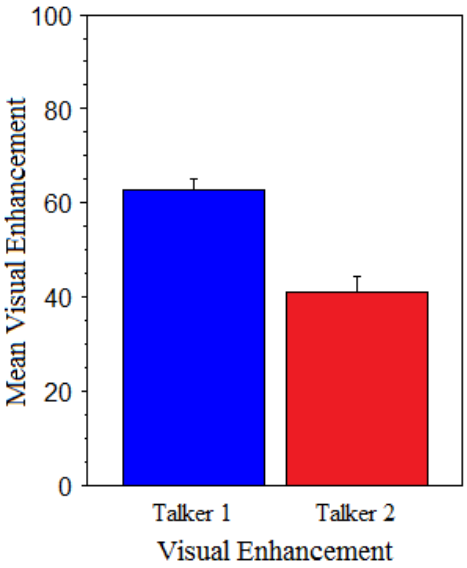

Figure 4. Mean visual enhancement scores for each talker

Measures of Integration

Results from the current study allow for the measure of audiovisual integration using two different methods: auditory enhancement (AE) and integration enhancement (IE). Both adopt a similar philosophy regarding the nature of audiovisual integration. It is the understanding that audiovisual speech perception is the result of the interaction between only three things: A-only input, V-only input, and a person's integration ability. This understanding is consistent with the theories of integration proposed by Grant and Seitz (1998), Sommers et al. (2005), and Tye-Murray et al. (2010). If this assertion is true, then integration ability can be assessed in AV performance after the ability to perceive the unimodal signals is accounted for. The following two calculations of integration are based on this approach.

Auditory Enhancement

Auditory enhancement can be described as the normalized amount of benefit resulting from the addition of an auditory speech signal to a visual speech signal (Sommers et al, 2005), where AE = (AV-V) / (1-V). Similar to VE, the amount of auditory enhancement is assessed by first subtracting the raw V-only score from the raw AV score. This measures the amount of non-normalized benefit attributed to adding the auditory signal to the visual signal. This difference is then divided by the result of subtracting the raw V-only score from 100 percent. The result is typically expressed as the percent of the potential improvement achieved when the auditory signal is added. In the current study, because performance in the A-only condition was controlled across participants and the formula normalizes for differences across participants in V-only ability, it is possible to use AE as a measure of integration ability. Figure 5 presents the resulting mean AE across participants for each talker. A repeated measures ANOVA indicated no difference in AE across the two talkers ($F(1, 22) = 0.2$; $p = .6319$). A scatter plot showing

AE of individual participants for each talker is shown in Figure 6. AE was not correlated across the two talkers (r = -.061).

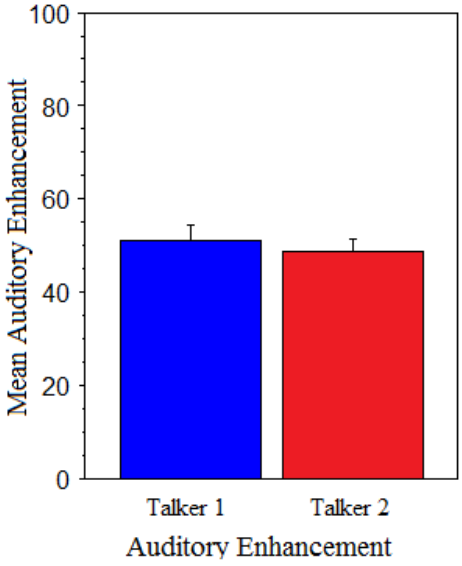

Figure 5. Mean auditory enhancement scores for each talker.

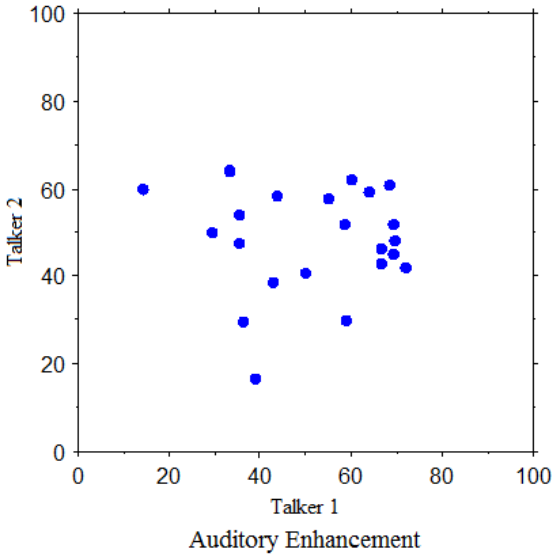

Figure 6. Scores of auditory enhancement for individual participants between talkers.

Integration Enhancement

Integration enhancement is described as the amount of normalized benefit that cannot be attributed to either auditory or visual performance: IE = (AVobserved – AVpredicted) / (1 – AVpredicted). To calculate IE, a prediction of AV performance is made based on probabilistic performance in each of the unimodal conditions. This prediction is calculated by multiplying the probability for error in the A-only condition by the probability for error in the V-only condition. The product subtracted from one is the predicted AV performance based on unimodal ability: AVpredicted = 1-[(1-A)*(1-V)]. The actual performance for the AV condition is then subtracted from the predicted AV performance. The resulting difference is a non-normalized index of integration ability because it accounts for performance in the AV condition that cannot be attributed to unimodal ability. The difference is then normalized by the amount of possible improvement that integration could have provided beyond unimodal performance. Figure 7

presents the mean IE values for each talker. A repeated measures ANOVA indicated no difference in IE across the two talkers ($F(1, 22) = 1.9$; $p = .6319$). A scatter plot showing IE of individual participants for each talker is shown in Figure 8. IE was not correlated across the two talkers ($r = -.051$).

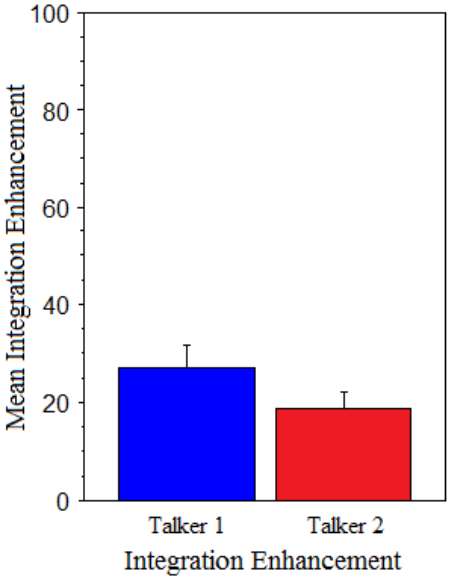

Figure 7. Mean integration enhancement scores for each talker.

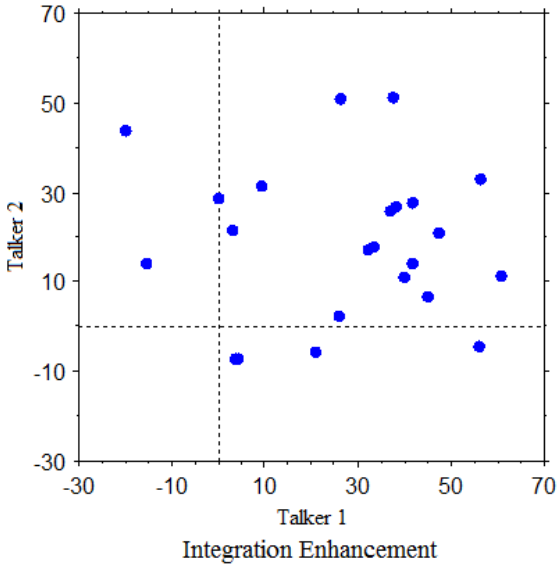

Figure 8. Scores of integration enhancement for individual participants between talkers.

Discussion

The current study was designed to investigate the influence of the degree of difficulty in the visual modality on audiovisual speech integration ability. As expected, a difference was seen in percent correct scores for the two talkers in the V-only condition. This finding indicates that the two talkers supplied differing levels of visual speech information to participants. The significant correlation between the V-only scores indicates that those who were good at lipreading Talker 1 were also good at lipreading Talker 2. Mean percent correct scores in the AV condition also showed significant differences between the two sets of scores, on average, collected with the two talkers. Because similar scores were obtained for each talker in the A-only condition and higher scores were obtained for Talker 1 versus Talker 2 in the V-only condition, the difference in AV scores was likely due to the difference in the amount of visual speech information gleaned from each talker. The V-only difference is also reflected in the difference noted for VE across the two talkers. The VE measure showed a difference in the benefit afforded by visual speech information when the auditory information was forced to very low performance (approximately 30% in the current study) for easy-to-lipread and hard-to-lipread talkers.

No significant differences were revealed between the two talkers for AE or for IE. At first glance, this lack of difference in the amount of integration suggests that changing the difficulty level for V-only input did not influence the level of integration achieved using the two types of signals, on average. However, it is also of note that a significant correlation was not seen for either AE or IE between talkers. This is clearly seen in Figures 6 and 8 where some participants demonstrated better AE and IE for Talker 1, the “good” visual signal, and some participants demonstrated better AE and IE for Talker 2, the “poor” visual signal. This latter

finding suggests that unimodal performance may indeed affect AE and IE, but that individuals are differentially affected when the degraded auditory speech signal is supplemented by a poor versus good visual speech signal.

The results of the comparisons across talkers for AE and IE do not support any of the theories proposed by the three models presented here. It is tempting to suggest that, because the average AE and IE scores were the same for the two talkers, results support a theory of audiovisual integration in which integration ability is thought to be constant regardless of the level of input. This would have been consistent with the assertions of Grant & Seitz (1998) who modeled audiovisual integration using a constant process mediating the integration of audio and visual inputs. However, because it was not possible to consistently predict the relationship between unimodal input and the level of integration that occurred, findings also do not support those that would be expected if either the PoIE (Lakatos et al., 2007) or any other model that would indicate that the level of the unimodal inputs affect the level of integration (Tye-Murray et al., 2010).

An alternative explanation for the current results must account for findings that indicate an unpredictable relationship between unimodal performance and integration ability. Integration ability *per se* would need to be described along with a factor or trait that allows the ability to be generalized across speakers. The extant literature regarding audiovisual integration often describes or implies that audiovisual speech perception can only be attributed to the combination of hearing, lipreading, and integration abilities. If that were true, the measures of integration presented here should be adequate for indexing the ability to integrate the two types of speech signals provided by Talker 1 and Talker 2. This is especially true when the same stimuli are presented in all three conditions (A-only, V-only and AV), as is the case here, because it controls

for variance in the complementariness and/or redundancy across the modalities that might occur if different words were presented across the test conditions. Results, like those reported here, that include a lack of correlation across talkers for AE and IE, are more consistent with a model that has more influencing factors than the three mentioned above. A review of the literature reveals that Grant & Seitz (1998) also found very little correlation between the ability to integrate sentence-level stimuli and nonsense syllables. The authors attributed the lack of association to differences in segmental cues, intonation and stress, contextual information available, and length of utterances between nonsense syllable and sentence stimuli, as well as differences in the way the brain processes these various types of information. None of these possible explanations could be applied here. Results would need to be replicated, but if the current account of the nature of integration persists, an alternative model of integration that allows for the introduction of more factors than the basic three is needed.

Limitations

One of the limitations of this study was the limited age range of participants. Results may not be able to be generalized to older adults and the elderly, especially if our ability to integrate changes as we age. A second limitation was the small number of males that were recruited and participated in this study; the inclusion of more male participants in future research could help with generalization across the two sexes. Third, the measures currently available may not accurately assess integration ability across multiple levels of unimodal input.

Clinical Implications

The findings from the present study support evidence from previous studies (Grant &

Seitz, 1998; Grant et al., 1998; Green & Kuhl, 1991; Sumby & Pollack, 1954; Summerfield, 1987), which showed that, when added to the auditory signal, visual speech cues could greatly enhance the listener's ability to understand speech. Aural rehabilitation programs should continue to include teaching of speechreading cues and lipreading to enhance speech perception ability of those with hearing loss. Describing the nature of audiovisual integration and the inputs that influence it will help rehabilitation experts assess and ultimately capitalize on realizing the fullest potential that could be provided by the audiovisual advantage.

Conclusions

The present study did not show significant differences in integration ability as the level (difficulty) of the unimodal input, specifically the visual input, varied. The results do not support any of the current theories of integration. The results do, however, support a new conceptualization of integration in which it is theorized that there are more contributing factors to integration than just the combination of the ability to hear, lipread, and integrate. Further study is also necessary to assess the implications of difficult unimodal inputs on integration ability as we age, with future research including a wider age range of participants.

References

American Speech-Language-Hearing Association, (1988). Guidelines for determining threshold level for speech. *ASHA, 30,* 85-89.

Grant, K. W. (2002, July). Measures of auditory-visual integration for speech understanding: A theoretical perspective [Letter to the editor]. *Journal of the Acoustical Society of America, 112*(1), 30-33.

Grant, K. W., & Seitz, P. F. (1998). Measures of auditory-visual integration in nonsense syllables and sentences. *Journal of the Acoustical Society of America, 104*, 2438-2450.

Grant, K. W., Tufts, J. B., & Greenberg, S. (2007). Integration efficiency for speech perception within and across modalities by normal-hearing and hearing-impaired individuals. *Journal of the Acoustical Society of America, 121*(2), 1164-1176.

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.

Green, K., & Kuhl, P. (1991). Integral processing of visual place and auditory voicing information during phonetic perception. *Journal of Experimental Psychology; Human Perception and Performance, 17*, 278-288.

Lakatos, P., Chen, C., O'Connell, M., Mills, A., & Schroeder, C. (2007). Neuronal oscillations and multisensory interactions in primary auditory cortex. *Neuron, 53*, 279-292.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Sommers, M., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear & Hearing, 26*(3), 263-275.

Sumby, W., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212-215.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audiovisual speech perception. In B. Dodd and R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3-51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., & Hale, S. (2010). Aging, audiovisual integration, and the principle of inverse effectiveness. *Ear & Hearing, 31*(5), 636-644.

Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., Hale, S., & Rose, N. (2008). Auditory-visual discourse comprehension by older and young adults in favorable and unfavorable conditions. *International Journal of Audiology, 47*(S2), S31-S37.

Tye-Murray, N., Spehar, B., Myerson, J., Hale, S., & Sommers, M. (Submitted). Reading your own lips: Common coding theory and visual speech perception. *Psychonomic Bulletin and Review.*