The Effects of Training on Intelligibility of Reduced Information Speech Stimuli

A Senior Honors Thesis

Presented in Partial Fulfillment of the Requirements for graduation *with distinction* in Speech and Hearing Science in the undergraduate colleges of
The Ohio State University

by

Kathryn James

The Ohio State University

June 2009

Project Advisor: Dr. Janet M. Weisenberger, Department of Speech and Hearing Science

Abstract

Although it is well known that individuals use visual cues to supplement hearing in noisy environments, research shows that we use visual cues even when the auditory signal alone is perfect. Speech perception is a multimodal process that uses both auditory and visual inputs on a continuous basis.

Recent research has focused on the benefit listeners receive from visual input when the auditory input is compromised in some way. One method of reducing information was initially used by Shannon and his colleagues, in which fine structure information was removed from speech syllables and replaced with band limited noise, while the temporal envelope was preserved. Results showed that stimuli were still identifiable even when these stimuli are reduced to two or three broad spectral bands.

However, other forms of auditory signal degradation, even though they yield good intelligibility in sentences, are much more difficult to identify as isolated syllables. One example is sine wave speech, in which the speech signal is reduced to a series of three sine waves that follows the formant frequency of the speech. Nonetheless, Exner showed that the intelligibility of sine wave speech syllables can show improvement with training. To date, no studies have been performed to determine whether stimuli similar to those employed by Shannon would also show improvements with training, or whether the level of performance observed reflects asymptotic intelligibility levels.

The present study evaluated the effects of training on performance with stimuli similar to those used by Shannon. To accomplish this, five listeners in the present study received 10 hours of auditory training with reduced-information speech syllables. Results showed significant improvement across training sessions for auditory-only performance, but a reduction in audio-visual integration after training. Implications for aural rehabilitation programs are discussed.

Acknowledgments

I would like to thank my advisor, Dr. Janet M. Weisenberger, for her support and guidance throughout this entire process. This opportunity has helped me to grow personally, professionally, and academically. I would also like to thank Michelle Hungerford for the assistance and guidance she provided in the lab. Lastly, I would like to thank my family and friends for their support through this process.

This project was supported by an ASC Undergraduate Scholarship and an SBS Undergraduate Research Scholarship.

Table of Contents

Chapter 1: Introduction and Literature Review

Although it is well known that individuals use visual cues to supplement hearing in noisy environments, research now shows that we use visual cues to supplement missing information in the auditory signal all the time. Auditory-visual integration occurs even when the auditory input alone is perfect.

A study by McGurk and MacDonald (1976) established the fact that people integrate visual and auditory stimuli even when it is not necessary to do so. McGurk and MacDonald tested 103 individuals with ages ranging from pre-school children to adults to see what the brain would perceive when conflicting auditory and visual stimuli were presented simultaneously. Results indicated that when the auditory stimulus [ba] was presented simultaneously with the visual stimulus [ga], the listeners responded with [da], suggesting that the brain fuses the two inputs together. This is what is now known as the McGurk effect.

In the example described above, the listener identifies cues to two different stimuli and in an effort to make sense of them fuses the stimuli, producing a result somewhere in the middle. Here, a bilabial auditory stimulus paired with a velar stimulus produces the response of an alveolar integrated stimulus, averaging the places of articulation. If the brain cannot correctly identify an explanation for the two discrepant inputs, then a combination response is produced. For example, one may perceive auditory [ka] and visual [pa] to be [kapa]. In this case the highly salient bilabial visual stimulus cannot be fused with the velar auditory stimulus input, and thus the response is a combination of the two inputs. The results of this study reiterate the importance of the visual input in speech integration.

Auditory Cues for Speech Perception

The auditory signal alone gives the listener a great deal of information about the signal. The auditory signal relays articulatory information about place, manner, and voicing for consonants. The place of articulation describes where the constriction of the air takes place in the mouth. The manner of articulation describes how the articulators make contact in order to produce the sounds. Voicing refers to the action of the vocal folds while the sound is being produced. In combination place, manner, and voicing help the listener differentiate between phonemes. The auditory information can be found in both the spectral and temporal aspects of a speech waveform.

Research has shown that much of the auditory signal is redundant, meaning that there is more information than is absolutely needed to identify the speech phoneme. Some of this redundant information can be removed and it should still be possible to identify the speech sound. One method of reducing information was initially used by Shannon and his colleagues (1995). Specifically, fine structure information was removed from the stimuli, and replaced with band limited noise, while the temporal envelope was preserved. Shannon assessed the effect of temporal envelope cues under reduced spectral conditions for the recognition of consonants, vowels, and sentences (Shannon, Zeng, & Wygonski, 1998). This study consisted of the following four experiments: the location of band divisions, warping the spectral distribution of envelope cues, shifting the frequencies of envelope cues, and finally spectral smearing. Results for experiment I indicate that the exact crossover frequencies in a four band representation are minor factors in speech recognition. Experiments II and III demonstrated results showing that warping the spectral distribution of envelope cues makes speech completely unintelligible, and that a tonotopic shift of the envelope pattern resulted in poor intelligibility as well. Experiments I and IV results show frequency divisions and overlap in carrier bands are not as critical for speech recognition and performance only diminished when the bands were broadly overlapping, smearing the tonotopic specificity of envelope cues (Shannon et al., 1998). Results further showed that stimuli are still somewhat identifiable even when these stimuli are reduced to two or three broad spectral bands. Shannon found that with 2 bands, auditory only performance for participants for single syllables that vary only in initial consonant was about 45% correct.

Visual Cues for Speech Perception

It is now known that we use visual information to supplement the missing information even when the auditory input alone is perfect. Visual cues are important; however, they can still be ambiguous (Jackson, 1988). The place of articulation can still be perceived to some degree. However, there are very few visual cues for manner of articulation and for voicing. For some phonemes it is easy to see the difference between them; for example, a bilabial /p/ will look very different from a velar /k/. However, a bilabial /p/ and a bilabial /b/ will be visually indistinguishable. Visually indistinguishable phonemes are referred to as visemes. Phonemes in a viseme group have identical places of articulation but differ in terms of manner of articulation and in voicing. In addition individual talker differences may aid in visual speech perception. The way in which a talker gestures and moves the head, eyes and mouth may provide cues to important information (Jackson, 1988).

Auditory-Visual Integration

Grant and Seitz (1998) found that for all but the most profoundly hearing impaired individuals, auditory–visual (AV) speech lends more accurate recognition than auditory (A) or visual (V) speech alone. They also found that among listeners, the degree to which people integrate auditory and visual stimuli varies greatly. Auditory only performance does not predict how well one can integrate, nor does visual only performance. Grant and Seitz argued that integration is a skill that is separate from either modality.

Auditory-Visual Integration Theories

A number of models have been proposed to describe speech integration across auditory and visual modalities. One important difference among these models is when integration occurs. The Pre-Labeling Model of Integration (PRE), as explained by Grant and Seitz (1998), combines information from the auditory and visual sources before identification. The PRE suggests that integration occurs early, preceding phoneme recognition (Grant, 2002). A prediction of auditory-visual recognition is made from auditory-only and visual-only confusion matrices. The perceived multimodal recognition for any speech sound should then be equal to or greater than the recognition from either the auditory or visual modality alone. The PRE can estimate integration efficiency, which is useful for developing rehabilitative programs which improve auditory-visual speech recognition.

The Fuzzy Logical Model of Perception (FLMP), constructed by Massaro (as cited in Grant, 2002), theorizes that separate decisions are made from each modality and integration results from a mediation of the two decisions. Like the PRE model, the FLMP also uses confusion matrices. Massaro (1998) suggests that incoming auditory, visual and auditory-visual information is independently assessed by listeners to extract summary descriptions of the incoming sensory information. The multiplicative integration rule used in determining auditory-visual speech perception performance in the FLMP is an optimal decision rule suggests Massaro (1987). It is applied to minimize the differences between obtained and predicted scores, and thus may be considered more likely to fit obtained bimodal scores rather than a prediction of optimal bimodal speech performance (Huffman, 2007). The FLMP assumes that the model predicts optimal integration. However, the problem with the FLMP is that human receivers often do better than predicted suggesting that this model may not be an accurate description of the process (Grant, 2002).

Recent Audiovisual Perception Studies

Recent studies in our laboratory by Huffman (2007) and Andrews (2007) have degraded stimuli in a manner similar to that used by Shannon et al. (1995). Huffman investigated the role of the auditory information in audiovisual speech integration. She found fairly good levels of identification performance even with just two broad spectral bands. Her results also suggest that systematically removing information from the auditory stimulus does not necessarily affect the degree of integration. Further exploration into reduced speech stimuli training was done by Exner (2008). Exner tested subjects using sine-wave stimuli, which is simply a different way of reducing the speech signal, with pre-training in audio-only and audio-visual conditions and then later tested post-testing to determine the effects of training on audio-visual integration with speech stimuli. Each participant was trained for two hours under auditory-only and auditory-visual conditions. Exner found that post-exposure testing yielded substantially better results. Parallel studies with increased exposure to stimuli similar to those used by Shannon et al. (1995)

have not been performed. The present study addressed this issue by evaluating the effects of training on performance with degraded stimuli using a method similar to Shannon et al. (1995). Auditory syllables were reduced to a waveform composed of a broadband noise fine structure that is modulated by the temporal envelope of the original speech stimulus recording. (Huffman, 2007). The study also investigates whether or not auditory-only training facilitates integration. In the study, five listeners received an initial pretest under auditory-only and audiovisual conditions without feedback, followed by five auditory-only training sessions with feedback. A mid-test under auditory-only and audiovisual conditions was administered without feedback, followed by five more auditory-only training sessions with feedback. A final posttest under auditory-only and audiovisual conditions was administered without feedback. Results were examined to evaluate the impact of training in both auditory only and auditory-visual conditions, to assess effects on auditory-visual integration.

Chapter 2: Method

Participants

Participants in this study included ten listeners and three talkers. Three females and two males, ages 16-22, participated in this study as listeners. Listeners were compensated $150.00 for their participation in the study. All reported having normal hearing. One female and two males participated as talkers. All three reported being native English speakers and were not compensated for their time.

Interfaces for Stimulus Presentation

Visual Presentation

Each participant was tested under the degraded auditory + visual condition, followed by training under degraded auditory only. For presentation of the visual portion of the stimulus, a 50 cm video monitor was positioned approximately 60 cm outside the window of the sound attenuating booth. The monitor was positioned at eye level, about 4 feet away from the participant's head. Stimuli were presented using recorded DVDs on a DVD player. For auditory only presentations the monitor was turned around.

Degraded Auditory Presentation

The degraded auditory stimuli were presented from the headphone output of the DVD player through Telephonics, 300-ohm TDH headphones.

Stimulus Selection

A set of eight CVC syllables were presented as stimuli for this study. The set of syllables was chosen to satisfy the following conditions:

1. Pairs of stimuli were minimal pairs, differing only in the initial consonant.
2. All stimuli contained the vowel /ae/, because it does not involve lip rounding or extension.
3. Multiple stimuli were used in each category of articulation, including; place (bilabial, alveolar), manner (stop, fricative, nasal), and voicing (voiced, voiceless).
4. All stimuli were presented without a carrier phrase.

Stimuli

The set of eight stimuli was used for both testing and training:

Bilabial: mat, bat, pat

Alveolar: sat, tat, zat

Velar: cat, gat

The four following dual-syllable (dubbed) stimuli were used in the degraded auditory + visual conditions. The first column represents the auditory stimulus, the second column indicated the visual stimulus.

bat-gat

gat-bat

pat-cat

cat-pat

Stimulus Recording and Editing

Each talker was recorded using a digital video camera. A microphone was used to record their voices directly into a computer, utilizing the software program Video Explosion Deluxe. This software allows the files to be stored in .wav format. All three of the talkers repeated the selected set of eight monosyllabic words five times. Then the speech samples were degraded using a MATLAB script designed by Delgutte (2003). The speech signal was filtered into two broad spectral bands. Then the fine structure was replaced with band-limited noise, while the temporal envelop remained intact. The result was a 2-channel stimulus, similar to those used by Shannon et al. (1998). Then the degraded auditory stimuli were dubbed onto the visual stimuli again using Video Explosion Deluxe.

Finally the software program Sonic MY DVD was used to burn the stimulus sets onto DVDs. 4 testing DVDs and 10 training DVDs were created for each talker. Each DVD contained stimuli in a random order to eliminate the possibility of memorization from the participants.

Procedure

Testing was conducted in The Ohio State University's Speech and Hearing Department. The participants were instructed by the examiner to read over the directions before the study began. The participants were tested in a sound attenuating booth and sat facing a 50 cm video monitor which was placed outside of the booth.

Due to testing errors two of the subjects received only the audio+visual testing condition. The results of the auditory-only and the visual-only conditions were left out of the results for these two listeners. These two participants were presented 12 DVDs, four for each talker, containing 60 randomly ordered stimulus syllables in the audio+visual condition only. The other three participants were administered a pre-test which consisted of 12 DVDs, each containing 60 randomly ordered stimulus syllables, four DVDs for each of the three talkers. The listeners were presented three DVDs each in the auditory-only and visual-only conditions, and then six DVDs in the audio+visual condition. Each DVD in the audio+visual condition used in testing included 30 stimuli with congruent auditory and visual components, and 30 stimuli with dissimilar auditory and visual components. The dissimilar stimuli were used to elicit McGurk-like responses. Participants were asked to listen/ and or watch each DVD. Participants were given a closed set of responses to choose from, and asked to verbally respond to what they perceived. No feedback was given.

Next, five auditory-only training sessions, each one hour long, were conducted. Each training session consisted of randomly selecting 6 of the 30 auditory-only training videos, two from each talker. The video monitor was then turned around to eliminate visual cues. Trial-by-trial feedback was provided to the participants on their answers. If the participant responded with an incorrect response the examiner would provide the correct answer. If the stimuli were correctly identified, then the examiner visually reinforced with the nod of a head.

After the five training sessions were completed a mid-test was administered, similar to the pretest. No feedback was given during the mid-test. Another five auditory-only training sessions were then conducted for each participant. Once again feedback was provided to the participant. After the final 5 sessions of training were completed, a final post-test was administered, without feedback. Testing and training took approximately 15-to-16 hours for each participant, and was broken up into two-to-three hour sessions with breaks to reduce fatigue.

Chapter 3: Results and Discussion

Results of the pre-test, mid-test, and post tests were analyzed to determine whether training affected identification performance in the auditory-only condition with degraded stimuli.

Analysis also dealt with whether training affected visual performance and integration ability. In addition to single-syllable stimuli, listeners were also presented with dual syllable (incongruent) stimuli, where the auditory stimulus differs from the visual stimulus. There is no "correct" response for dual syllable presentations. However, the responses can be categorized according to whether the listener chose a response corresponding to the auditory or visual stimulus, or chose some other response, which matched neither the auditory nor visual stimulus used in the incongruent pairing.

Percent Correct Performance

Figure 1 shows the overall percent correct performance of listeners for the Auditory-only (A), Visual-only (V), and Audio+visual (A+V) conditions for the pre-test, mid-test, and post-test, averaged across talkers and listeners. Improvement in the auditory-only condition was observed, as well as an improvement in the audio+visual condition from pre-test to post-test. The visual-only condition did not show such improvements, as performance for this condition stayed relatively the same. This makes sense because listeners were not trained in the visual-only condition. The auditory-only performance progressively increased from pre-test to mid-test to post-test. This implies that that auditory-only training did improve the listener's ability to identify the degraded stimuli. A two-factor, repeated-measures ANOVA showed a significant main effect of test (pre, mid, post), ($F(2, 8)=51.2$, $p<.001$). A significant main effect of modality was also observed, ($F(2,8)=197.79$, $p<.001$). Finally, confidence intervals indicated significant differences in all means contrasts. A significant interaction effect (test x modality) was also found, ($F(4,16)=10.11$, $p=.003$).

Figure 2 shows auditory-only responses for each talker in the auditory-only pre-test and post-test, averaged across listeners. Intelligibility of all three talkers improved with training, with talker KS showing the largest improvement. The data show only small variability in the percent correct for the post-test among the three talkers, meaning that there were no important talker differences in the auditory-only condition.

Figure 3 shows visual-only responses for each talker in the visual-only pre-test and post-test, averaged across listeners. Talkers DA and JK showed improved intelligibility, with talker KS actually showing a slight decrease in intelligibility. This is noteworthy because KS showed the largest improvement in the auditory-only condition.

Figure 4 shows audio+visual response for each talker in the audio+visual pre-test, and post-test, averaged across listeners. Once again, talkers DA and JK show very similar amounts of improvement with KS showing the greatest improvement.

Figure 5 shows overall percent correct responses for auditory-only training, for each listener, averaged across talkers. For all listeners, performance improved across sessions. Listener HG had higher performance levels overall, while listener SJ showed the greatest improvement throughout training sessions. All listeners showed more gains in stimulus identification in the first five sessions.

Tables 1 and 2 show confusion matrices for auditory-only performance, to give an indication of how individual stimuli were perceived before and after training. Every stimulus became more intelligible from pre-test to post-test.” Mat” was the most correctly identified stimulus in the pre-test with a 60% correct identification, and improving to 92% in the post-test. Stimulus “sat” was the most correctly identified in the post-test with 94% correct identification.

Tables 3 and 4 show confusion matrices for audio+visual performance, to investigate whether training affected audio+visual integration of individual stimuli. Once again every stimulus improved from pre-test to post-test. Correct identification was overall better than for the auditory-only condition. The stimulus “bat” was most commonly confused with “pat” or “mat”, which is not surprising because these three stimuli are in the same viseme category. “Cat” and “gat” were commonly mistaken for each other, also influenced by their viseme category which is velar. “Bat” was often mistaken for “zat”. “Bat” is a bilabial and “zat” is an alveolar in place of articulation thus, the confusion indicated the use of voicing cues and not place cues. One final syllable to note is “tat”. In the pre-test tat was repeatedly mistaken for “cat”. This also indicates the use of voicing cues, as both are voiceless stops. “Tat” is interesting, as it was the most ambiguous syllable, resulting in mistakes across all places of articulation, more so than any other syllable.

Figure 6 shows the amount of audio+visual integration, where integration is defined as the difference between audio+visual performance and the better single modality, (A), or (V), averaged across talkers and listeners. Results show a decrease in integration from pre-test, to mid-test, to post-test, suggesting that benefits from training improved only the (A) performance. These data support Grant and Seitz’ findings in that audio+visual speech lend more accurate recognition than either modality alone. Integration is a separate ability and thus must be trained in addition to both auditory-only and visual-only training.

Integration of Discrepant Stimuli

Figure 7 shows overall percent dual-syllable responses for the pre-test, mid-test, and post-test, averaged across talkers and listeners. Listeners showed increased reliance on the auditory modality after the first five hours of training. As mentioned there are no correct responses, for the stimuli, and thus responses are shown as the percent of the time the listener chose the auditory stimulus, the visual stimulus, or some other stimulus that might indicate integration. From the pre-test to mid-test, the percentage of the “auditory” responses increased dramatically, indicating that training in the auditory-only condition, increased reliance on the auditory modality. The percentage of auditory responses stayed about the same from mid-test to post-test. Reliance on the visual stimulus showed no changes across tests, while the percentage of “other” responses decreased from pre-test to post-test.

Figure 8 shows percent McGurk-type integration for dual-syllable responses, averaged across talkers and listeners. Listeners showed a higher percentage of combination responses before training and more fusion responses after training. Listener’s recognized that they were not being presented with combination stimuli in training and thus ruled out these responses.

Chapter 4: Summary and Conclusion

Overall, the results of this study indicate that training with reduced information auditory stimuli does result in performance improvements. All subjects showed substantial improvement across training sessions. However, the decrease in audio-visual integration observed from pre-test to post-test indicates that auditory-only training does not improve the integration process. Results confirm the fact that auditory-only performance does not predict how well one can integrate, nor does visual-only performance (Grant and Seitz, 1998). The present results support Grant and Seitz's argument that integration is a skill that is separate from either modality.

Training under auditory-only conditions is not sufficient to maximize benefits from aural rehabilitation. An effective aural rehabilitation program should contain training under dual-modality conditions as well as modality-specific training to maximize speech perception for hearing impaired persons.

Future studies should investigate training on the audio+visual condition since integration is an independent skill. Modality specific training does not appear to improve integration and so a study needs to be performed to investigate whether or not dual-modality training is the best way to maximize benefits for aural rehabilitation.

Chapter 5: References

Andrews, B. (2007). *Auditory and visual information facilitating speech integration.* Senior Honors Thesis, The Ohio State University

Delgutte B. Mathematical models of cochlear nucleus onset neurons. I. Point neuron with many weak synaptic inputs. *J. Comput. Neurosci.* 2003; **14**:71-90

Exner, M. (2008) Training *effects in audio-visual Integration of sine wave speech.* Senior Honors Thesis, The Ohio State University.

Grant, K.W. & Seitz, P.F. (1998). Measures of auditory-visual integration in nonsense syllables and sentences. *The Journal of the Acoustical Society of America, 104 (4),* 2438-2450.

Huffman, C. (2007). *The role of auditory information in audiovisual speech integration.* Senior Honors Thesis, The Ohio State University.

Jackson, P.L. (1988). The theoretical minimal unit for visual speech perception: Visemes and coarticulation. *The Volta Review, 90 (5)*, 99-114.,

Massaro D. (1998) *Illusions and issues in bimodal speech perception.* in Auditory-Visual Speech Processing Conference. Terrigal, Sydney, Australia. p. 21-26

Massaro, D.W. (1987). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry (*Erlbaum, Hillside, NJ).

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

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270,* 303-304.

Shannon, R.V., F.G., Wygonski, J. (1998). Speech recognition with altered spectral distribution of envelope cues. *The Journal of the Acoustical Society of America*, 104 (4), 2467-2475.

Table 1: Pre-test Audio-only

Response

Stimulus	bat	pat	mat	gat	cat	zat	tat	sat	dat	nat	pcat	ptat	bgat	bdat
bat	**32%**	16%	7%	6%	6%	6%	4%	0%	7%	4%	3%	2%	4%	0%
pat	17%	**23%**	3%	10%	15%	5%	7%	1%	9%	1%	1%	2%	2%	1%
mat	12%	5%	**60%**	5%	2%	5%	2%	0%	0%	7%	0%	0%	2%	0%
gat	28%	8%	6%	**19%**	9%	5%	1%	4%	12%	3%	0%	0%	4%	2%
cat	1%	13%	3%	9%	**41%**	2%	3%	1%	11%	1%	2%	5%	7%	0%
zat	41%	9%	6%	3%	3%	**22%**	3%	6%	0%	3%	0%	0%	0%	3%
tat	14%	20%	9%	0%	6%	0%	**17%**	3%	9%	6%	3%	6%	6%	3%
sat	8%	0%	3%	3%	10%	13%	0%	**50%**	0%	5%	3%	3%	3%	3%

Table 2: Post-test Audio-only

Response

Stimulus	bat	pat	mat	gat	cat	zat	tat	sat	dat	nat	pcat	ptat	bgat	bdat
bat	**67%**	15%	2%	16%	2%	1%	2%	0%	0%	0%	0%	0%	0%	0%
pat	12%	**57%**	4%	5%	6%	1%	15%	0%	0%	0%	0%	0%	0%	0%
mat	8%	0%	**92%**	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
gat	25%	6%	1%	**57%**	6%	1%	4%	0%	0%	0%	0%	0%	0%	0%
cat	0%	8%	0%	6%	**61%**	0%	3%	0%	0%	0%	0%	0%	0%	0%
zat	41%	0%	0%	0%	0%	**59%**	0%	0%	0%	0%	0%	0%	0%	0%
tat	0%	16%	0%	5%	16%	0%	**59%**	3%	0%	0%	0%	0%	0%	0%
sat	0%	0%	0%	0%	0%	6%	0%	**94%**	0%	0%	0%	0%	0%	0%

Figure 3: Pre-test Audio+Visual

Response

Stimulus	bat	pat	mat	gat	cat	zat	tat	sat	dat	nat	pcat	ptat	bgat	bdat
bat	**61%**	23%	11%	3%	0%	1%	1%	0%	1%	0%	0%	0%	0%	0%
pat	28%	**53%**	5%	2%	1%	1%	2%	1%	4%	2%	0%	2%	0%	1%
mat	11%	10%	**75%**	0%	1%	1%	0%	0%	1%	1%	1%	0%	0%	1%
gat	1%	1%	0%	**57%**	20%	0%	5%	1%	11%	4%	0%	0%	0%	0%
cat	0%	1%	0%	22%	**62%**	1%	8%	0%	3%	1%	2%	0%	0%	0%
zat	13%	2%	1%	5%	0%	**53%**	4%	10%	7%	3%	0%	1%	1%	0%
tat	6%	4%	0%	10%	25%	2%	**29%**	5%	15%	1%	0%	2%	1%	0%
sat	0%	1%	0%	1%	1%	9%	6%	**79%**	1%	0%	0%	0%	1%	0%

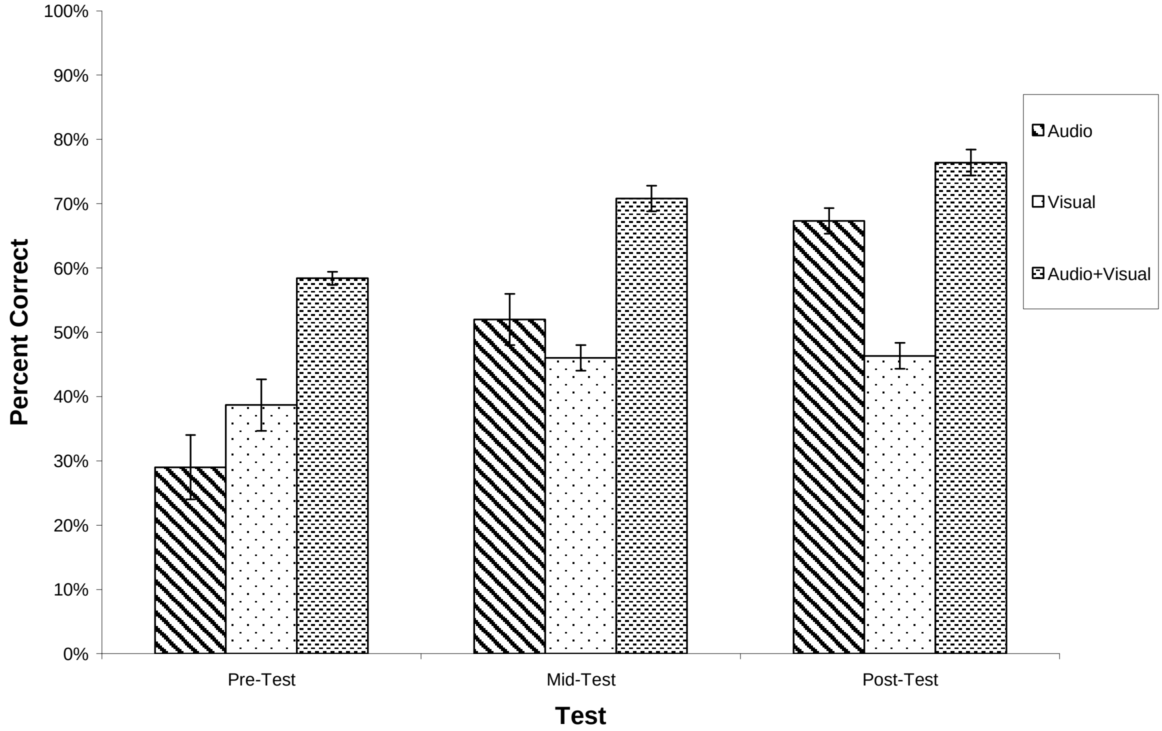
Figure 4: Post-test Audio+Visual

Response

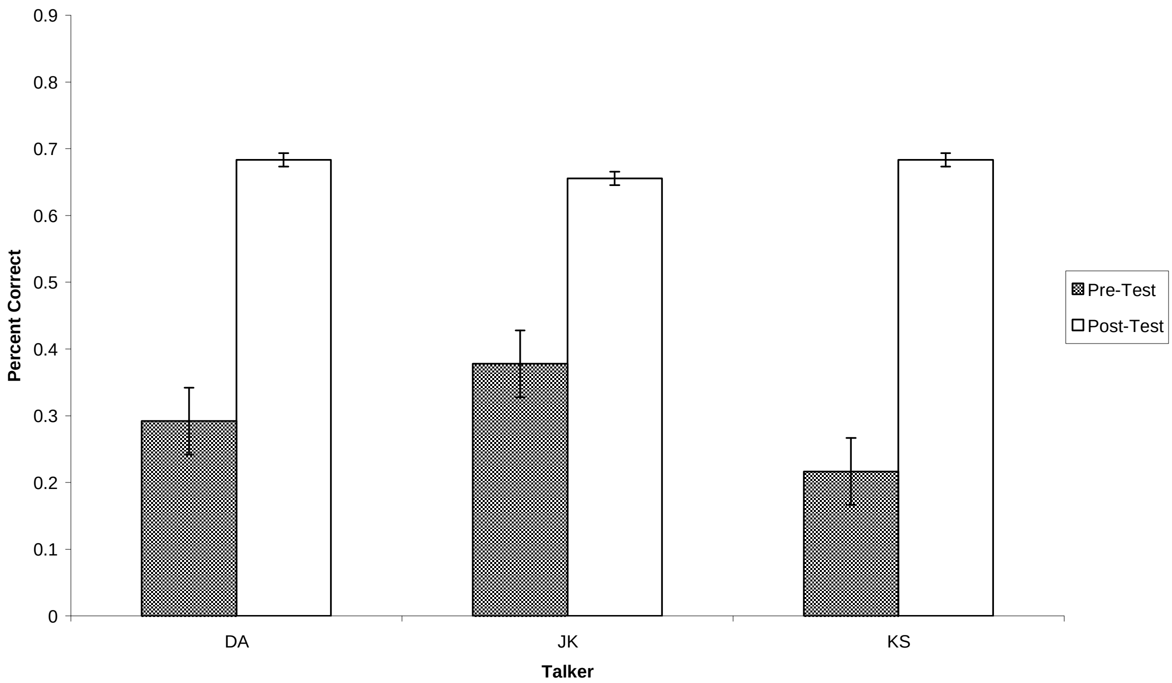
Stimulus	bat	pat	mat	gat	cat	zat	tat	sat	dat	nat	pcat	ptat	bgat	bdat
bat	**67%**	19%	3%	3%	0%	2%	5%	0%	0%	0%	0%	0%	0%	0%
pat	21%	**67%**	0%	2%	2%	1%	7%	0%	0%	0%	0%	0%	0%	0%
mat	6%	6%	**88%**	0%	0%	1%	1%	0%	0%	0%	0%	0%	0%	0%
gat	1%	1%	0%	**73%**	20%	0%	3%	0%	1%	0%	0%	0%	0%	0%
cat	1%	0%	0%	11%	**84%**	0%	5%	0%	0%	0%	0%	0%	0%	0%
zat	27%	1%	1%	0%	0%	**67%**	1%	3%	0%	0%	0%	0%	0%	0%
tat	3%	17%	0%	5%	10%	0%	**60%**	1%	3%	1%	0%	0%	0%	0%
sat	0%	0%	0%	0%	1%	3%	1%	**96%**	0%	0%	0%	0%	0%	0%

List of Figures

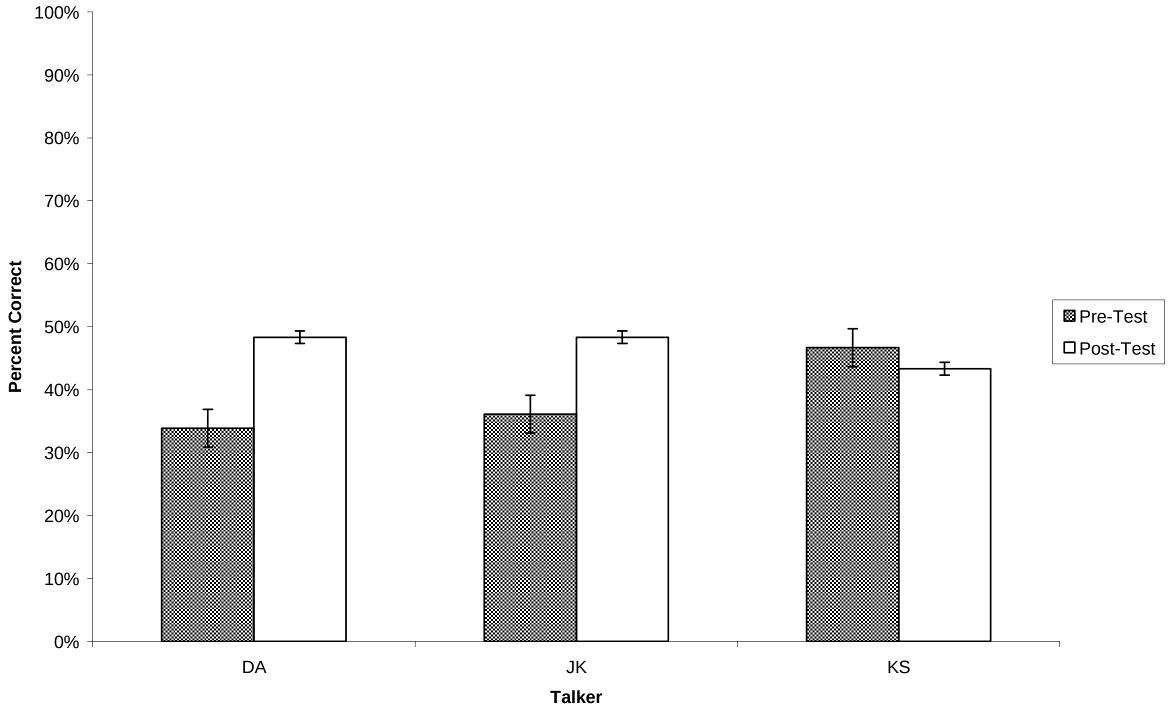
Percent Correct Responses Across Tests
Figure 1



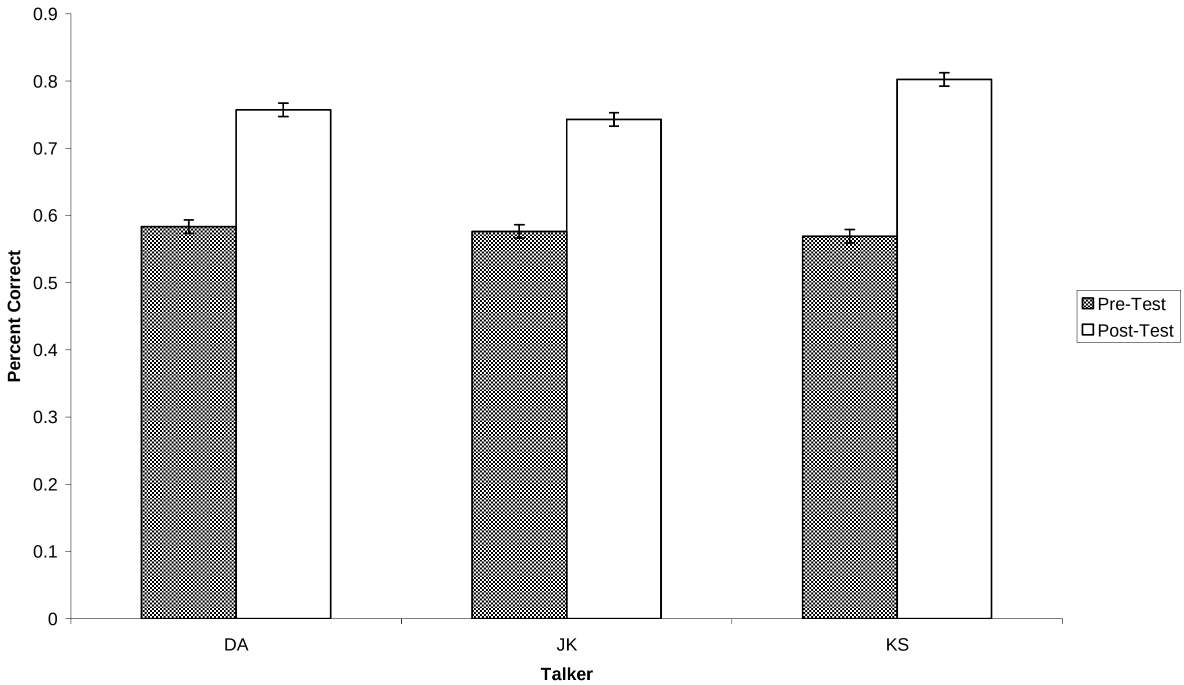
Auditory Only Responses
Figure 2



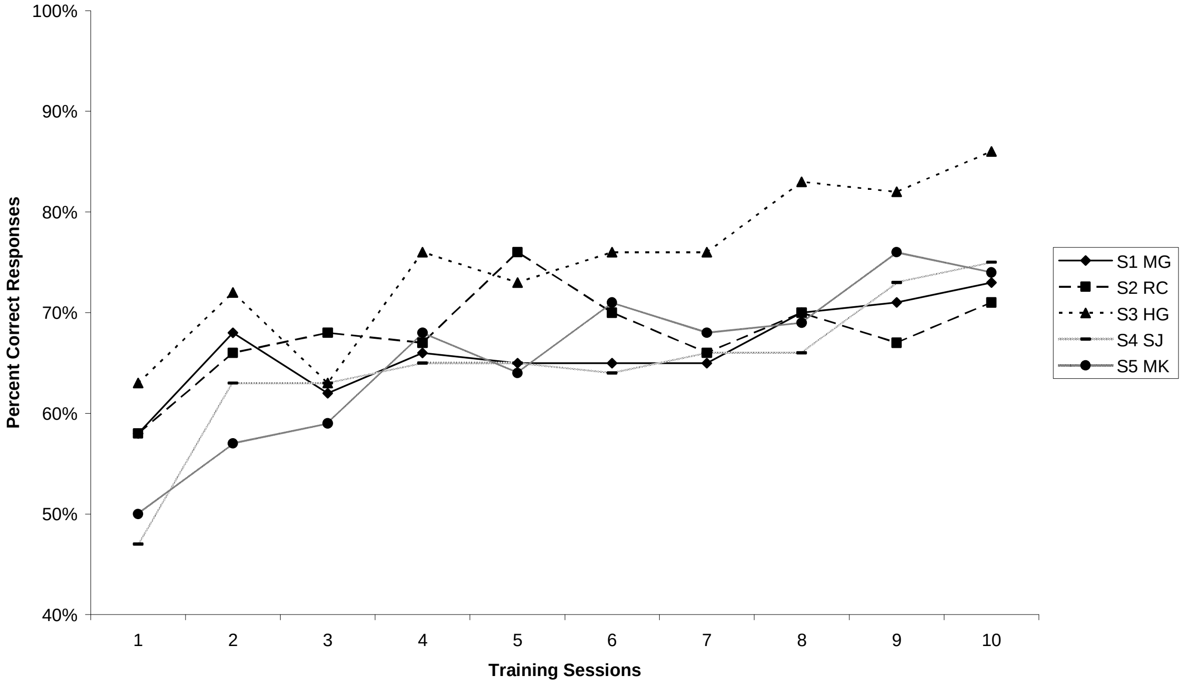
Visual Only Responses
Figure 3



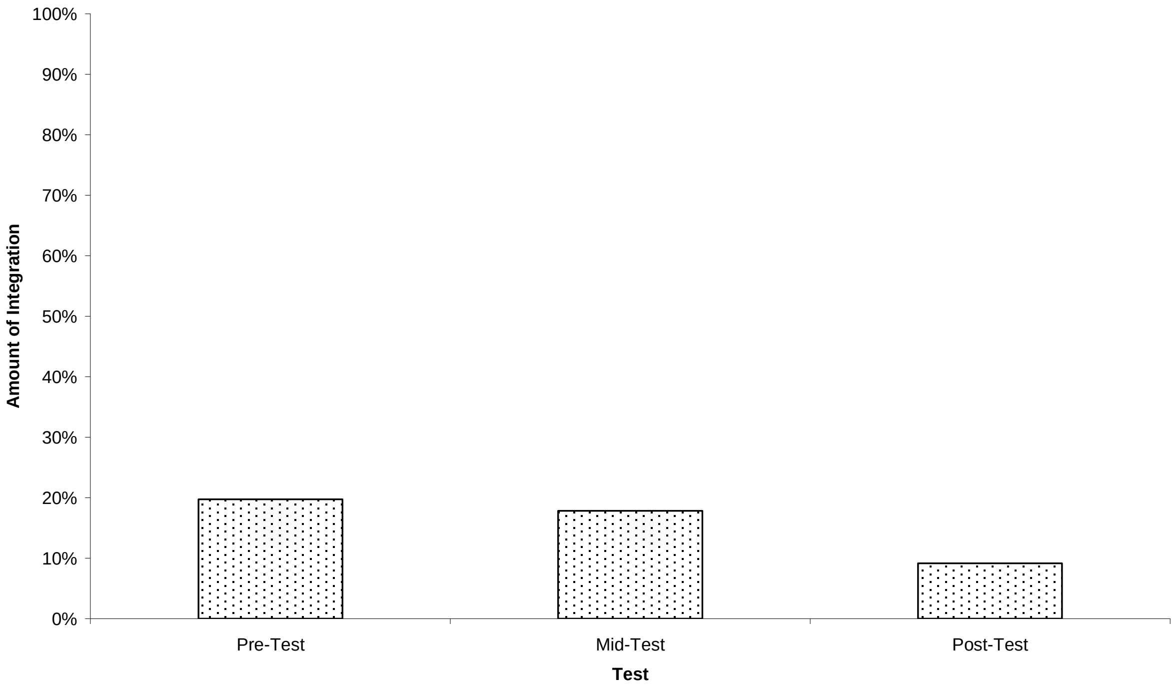
Auditory+Visual Responses
Figure 4



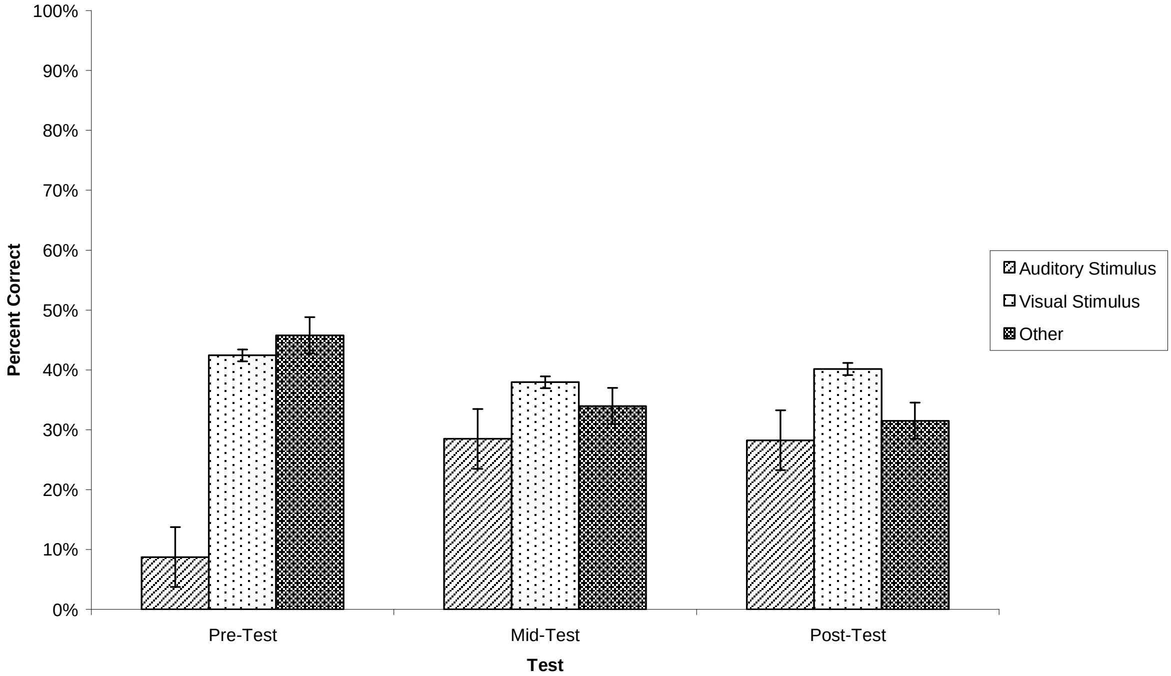
Percent Correct Responses for Training Sessions Across all Talkers
Figure 5



Amount of Integration
Figure 6



Overall Percent Dual-Syllable Response Rate
Figure 7



McGurk Type Integration
Figure 8