Can Audio-Visual Integration Improve with Training?

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

Sara DiStefano

The Ohio State University
June 2010

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

Abstract

The integration of visual cues and auditory speech cues is a process used by listeners in both normal and compromised listening situations. Audio+visual integration of speech appears to be independent of the ability to process auditory-only or visual-only speech cues. Grant and Seitz (1998) argued for independence of this process based on the fact that integration could not be easily predicted by auditory-alone or visual-alone performance. Gariety (2009) and James (2009) provided additional support for this argument. In their studies, training on degraded auditory speech syllables under auditory-only conditions improved auditory performance but not audio+visual performance.

The question remains whether integration itself is an ability that can be trained. In the present study, five listeners received ten training sessions in the audio+visual condition with degraded speech syllables similar to those used by Shannon et al. (1995). A comparison of pre-training to post-training scores showed little to no improvement in auditory-only and visual-only identification, but a substantial improvement in audio+visual performance. These results provide further support for the idea that integration is an independent process, and argue for the incorporation of audio+visual integration training into aural rehabilitation programs.

Acknowledgments

I would like to thank my advisor, Dr. Janet M. Weisenberger, for providing me with this wonderful opportunity to work with her on this thesis. Through her guidance and support I was able to gain experience and grow academically. I would also like to thank my subjects for their flexibility and commitment throughout this thesis project.

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

Table of Contents

Chapter 1: Introduction and Literature Review

It has long been known that when an auditory signal is compromised in some way, such as in a hearing loss or a noisy environment, visual cues are then used to compensate and aid in speech perception. Visual cues enhance features of the auditory signal and supplement missing auditory information to aid in the perception of speech under such conditions. However, research by McGurk and MacDonald (1976) gives evidence that even individuals with normal hearing, in situations in which the auditory signal is completely intelligible, also use visual cues to enhance their speech perception.

In their experiment, the presentation of an auditory syllable such as a bilabial /ba/ together with a visual, velar consonant /ga/ resulted in the perception of the alveolar consonant /da/, which is a fusion of /ba/ and /ga/. This is the brain's attempt to average the place of articulation of the two. Reversing this process, an auditory /ga/ with a visual /ba/, resulted in the response, /baga/. This combination response is due to the influence of the strong visual bilabial stimulus and the brain's inability to explain the discrepant inputs. This phenomenon, known as the McGurk effect, indicated that a visual stimulus can actually change the perception of an auditory sound. Today it is understood that audio-visual integration is an automatic process that occurs unconsciously. An understanding of the audio-visual integration process requires a knowledge of the processes underlying unimodal auditory speech perception and visual speech perception.

Auditory Cues for Speech Perception

The auditory signal provides three main cues for identifying consonants. These include place of articulation, manner of articulation and voicing. Place of articulation refers to where in the mouth the sound is produced, or the location of constriction in the oral cavity. These locations consist of bilabials, labiodentals, interdentals, alveolars, palatal-alveolars, palatals and velars. The manner of articulation describes how the articulators come in contact with one another to form a sound. Manners include stops, fricatives, affricates, liquids and glides. The third cue is voicing, which refers to the presence or absence of vocal fold vibrations. If the vocal folds are vibrating during speech production the sound is said to be voiced, whereas if the vocal folds are not vibrating during the production the sound is voiceless. A characteristic of stop consonants embedded with voicing is voice onset time (VOT). Voice onset time is defined as the length of time that passes between when a stop consonant (/p,t,k,b,d,g/) is produced and when voicing occurs. All of this information is incorporated in the spectral and temporal envelopes of the speech waveform (Ladefoged, 2006).

Visual Cues for Speech Perception

As shown by McGurk and MacDonald, visual inputs also serve as important cues for speech perception. However, less information can be obtained from visual cues. Place of articulation is essentially the only observable feature, and that in itself can even be ambiguous (Jackson, 1988). Since there is only minimal information on manner of articulation and none concerning voicing, it is difficult to correctly identify a sound from

visual information alone.

This difficulty is a result of groups that are referred to as visual phonemes, or visemes (Jackson, 1988). Items in a viseme group have an identical place of articulation, but differ in manner and voicing. For example, the bilabials /p, b, m/ constitute a viseme group. Difficulty reading speech when there is no auditory signal also occurs due to characteristics of individual talkers. A study done by Jackson found that talkers who created more viseme categories were easier to speechread, compared with those who produced fewer. Other cues provided by the talker that may aid in speech perception include gestures and movements of the eyes, head and mouth. These cues are also helpful in situations involving a degraded auditory signal.

Speech Perception with Reduced Auditory and Visual Signals

Speech is still highly perceptible even in situations with a reduced auditory signal, due in part to the amount of redundant information that is provided in these signals. A study by Shannon and colleagues (1995) found that acoustic speech waveforms contained more information than absolutely necessary to identify a speech sound. They found that replacing the fine-structure information of syllables with band-limited noise, while preserving the temporal envelope, produced sounds that are still highly identifiable. Identification improved as the number of noise-bands increased, but high levels of speech recognition could be reached even with only three bands of modulated noise (Shannon et al., 1995). In 1998, Shannon and his colleagues expanded the previous study by conducting four experiments that explored which parameters of a reduced signal are most

critical for speech recognition. They found that the exact cutoff frequencies which define the three bands and that the selectivity of the envelope carrier bands were not critical for speech recognition. On the other hand, warping the spectral distribution of envelope cues and shifting the tonotopic organization of the envelope pattern resulted in poor intelligibility of speech (Shannon et al., 1998). This suggests that the temporal envelope cue is a component of a sound which gives it its defining characteristics.

A study by Remez et al. (1981) also focused on speech intelligibility under degraded auditory conditions; however, this study reduced speech sounds to three sine waves representing the first three formants of the original signal. This method of degrading an auditory signal still yielded high speech intelligibility levels, even though listeners perceived these sounds as unnatural (Remez et al., 1981). This finding, like the work of Shannon et al., suggests that speech can still be intelligible even when substantial amounts of information have been removed.

As seen in the McGurk effect, visual cues are also taken into account during speech perception. Like auditory input, visual cues do not have to be perfect to enhance speech perception. Munhall et al. (2004) found that auditory speech intelligibility levels were increased by adding information from visual images that had been degraded through band-pass and low-pass filtering. Results also indicated that even a limited spatial frequency spectrum is sufficient for audio+visual speech perception (Munhall et al., 2004).

Audio-Visual Integration of Reduced Information Stimuli

The study of audio-visual integration in hearing-impaired persons is especially

important, as it constitutes a case of visual input combined with reduced auditory signals. Grant and Seitz (1998) assessed integration abilities across hearing-impaired listeners using a variety of auditory-visual integration measures to establish whether integration is a process that is independent of auditory-only or visual-only processing. Congruent and discrepant nonsense syllables were degraded using a bandpass filterbank with four nonoverlapping filter bands between 300 and 6000 Hz. Congruent stimuli are described as having the auditory signal "match", or be in synchrony with, the visual articulators. Discrepant stimuli on the other hand are described as having the auditory signal and visual cue "out of sync". These stimuli can either be misaligned or have another auditory signal dubbed on to a different visual cue. These degraded syllables were then presented to listeners in the auditory (A), visual (V), and audio+visual (A+V) conditions. Results showed that even with an extremely reduced auditory signal, AV benefit was still significantly high. However, because a person's audio-only or visual-only performance could not predict their integration efficiency, Grant argued that audio+visual integration is independent of a person's ability to extract auditory and visual information from speech. Results concerning integration independent of auditory or visual cues, however, showed little association between integration measures derived from nonsense syllable tests and those derived from sentence tests (Grant and Seitz 1998).

Previous studies in our laboratory have also used degraded signals in studying audio+visual integration. For example, Feleppelle (2008) examined the role of the auditory signal in audio+visual integration to determine whether the amount of information reduction in the auditory signal is a contributor to the large degree of

variability observed in the audio+visual integration benefit achieved by listeners. This study also took talker variability into account. Listeners were tested on their speech perception abilities in auditory-only, visual-only, and audio+visual conditions. Four levels of auditory degradation using a method similar to that previously described by Shannon et al. (1998) were tested. The stimuli were degraded using 2, 4, 6, and 8 bandpass filter channels. Spectral fine structure was removed and replaced with noise, while temporal envelopes were preserved. Results from this study indicated that listeners were able to integrate audio and visual cues even when there was considerable information missing from the auditory signal. Results also showed that while increasing degradation of the auditory information negatively affected speech perception performance in the A and AV conditions, the amount of AV benefit, defined as the difference between AV and the best single modality, remained relatively consistent.

Andrews (2007) examined the AV integration benefit produced by fourteen different talkers using the stimuli above. Results suggested that talker characteristics may play a major role, given the significant variability in the auditory-only and audio+visual conditions that was observed across talkers. Also, it was found that the performance produced with a talker in the auditory-only or visual-only condition is not a predictor of the amount of audio+visual integration that a talker is able to produce. For example, the talkers producing the most audio+visual integration were not those with the highest auditory-only or visual-only intelligibility, supporting Grant's argument that audio+visual integration is a process independent of audio-alone or visual-alone processing.

Studies in our laboratory have also employed degraded auditory stimuli similar to

those used by Remez et al. (1981). Tamosuinas (2007) degraded congruent and discrepant speech syllables using four different sine wave reductions (F0, F1, F2, and F0+F1+F2) and presented these signals to listeners in A, V, and A+V conditions. For both types of stimuli, results showed very low auditory-only and audio+visual performance, and little evidence of integration. Visual scores were consistent with previous studies (about 30% correct). This suggested that sinewave speech, at least in individual syllables, is too degraded a signal to facilitate auditory-visual integration (Tamosuinas 2007).

Effects of Training in Recent Studies

The low levels of performance observed in the studies with sine wave stimuli above led to questions about whether the lack of familiarity of these stimuli might have influenced performance. This issue was addressed in three subsequent studies. Exner (2008) explored causes for the lack of integration and benefit seen in Tamosiunas' (2007) study. Listeners in Exner's study were provided with two hours of auditory training in sine wave speech perception to see whether the results of Tamosiunas' study were a product of unfamiliarity with sine wave speech or whether the auditory signal was degraded past the point of identification. Exner found that training with highly degraded auditory stimuli can lead to improvements in intelligibility. However, these benefits were confined to the auditory-only condition, and the amount of integration did not show a significant change as a function of training. Longer as well as separate training sessions for integration and auditory listening tasks were suggested to increase integration skills.

Gariety (2009) investigated whether longer training periods under the auditory

condition would improve the amount of audio+visual integration across listeners. Although the amount of training was increased from two to ten hours and significant improvement in auditory-only performance across training sessions was seen, the amount of audio+visual integration still did not change (Gariety, 2009). A similar study by James (2009) tested whether training in the auditory condition improved performance with degraded stimuli similar to those used by Feleppelle (2007). Auditory-only performance improved across ten training sessions, but again integration ability was unaffected. These studies suggest that training in the audio+visual condition might be necessary for improving audio+visual integration scores.

Present Study

Although the work cited above indicates that auditory-only performance can benefit from training, the question remains whether audio+visual integration is an ability that can improve with training. The present study addressed this question by providing ten hours of audio+visual training to normal-hearing participants. Degraded auditory input was paired with visual stimuli to determine whether audio+visual integration scores improve across pre-, mid-, and post-tests. The auditory stimuli were digitally-recorded monosyllables differing only in initial consonant, and were degraded in a manner similar to that of Shannon et al (1998). It was hypothesized that audio+visual integration scores would improve after listeners completed the training sessions. However, if integration truly is independent of single-modality performance, no improvements in audio-only or visual-only conditions should be observed. Results from this study should provide some insights for the development of effective aural rehabilitation programs for persons with

hearing impairments. These individuals can be trained to maximize their audio+visual integration benefit to overcome the challenge of auditory system damage paired with a difficult listening situation.

Method

Participants

Participants in the present study included five listeners. Four males and one female, ages 20-22 years, participated. All five reported having normal hearing as well as normal or corrected vision. None of the participants had a background in Speech and Hearing Science. Participants were compensated $90 for their participation. Materials previously recorded from five adult talkers, two males and three females, were used as stimuli.

Stimuli Selection

A limited set of eight syllables were presented as stimuli for the study. All syllables satisfied the following conditions:

1. The pairs of stimuli were minimal pairs; they differed only in the initial consonant.
2. All stimuli contained the vowel /ae/, used because it does not involve lip rounding or lip extension, which can create speech reading difficulties.
3. Multiple stimuli were used in each category of articulation, including: place (bilabial, alveolar, velar), manner (stop, fricative, nasal), and voicing (voice, unvoiced).
4. All were presented without a carrier phrase.

Stimuli

For each of the conditions the same set of single-syllable stimuli were used:

Bilabial: bat, mat, pat

Alveolar: sat, tat, zat

Velar: cat, gat

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

bat-gat

gat-bat

pat-cat

cat-pat

Stimuli Recording and Editing

Stimuli from recent studies (e.g., James, 2009) were used in this experiment to permit comparisons of results. Speech samples from five talkers 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 envelope 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 using Video Explosion Deluxe, a commercial video editing program.

Finally the software program Sonic MY DVD was used to burn the stimulus sets onto DVDs. Four DVDS were created for each talkers. Each DVD contained sixty stimuli in a random order to eliminate the possibility of memorization from the participants.

Visual Presentation

Each participant was pre-tested under degraded auditory, visual, and audio+visual conditions, followed by training with degraded audio+visual presentation. For presentation of the visual portion of the stimulus, a 50 cm video monitor was positioned approximately 60 cm outside the window of a sound attenuating booth. The monitor was positioned at eye level, about 120 cm away from the participant seated inside the booth. Stimuli were presented using recorded DVDs on a DVD player. For auditory only presentation the monitor screen was darkened.

Degraded Auditory Presentation

The degraded auditory stimuli were presented from the headphone output of the DVD player through 300-ohm TDH-39 headphones at a level of approximately 75 dB SPL.

Testing Procedure

Testing was conducted in The Ohio State University's Speech and Hearing Department in Pressey Hall. Participants were instructed to read over a set of instructions explaining the procedure and listing a closed-set of response possibilities. The response set also included options that might reflect McGurk-type fusion or combination responses for the discrepant stimuli.

Participants were individually tested in a sound attenuating booth facing a video monitor placed outside the booth. Auditory stimuli were transmitted through headphones inside the booth. The examiner recorded and scored the participant's responses through

an intercom system. Each participant was administered a pre-test using stimuli selected from a set of 15 DVDs, each containing 60 randomly ordered syllables, three DVDs for each of the five talkers. In the pretest, the listeners were presented with one DVD from each talker in each of three listening conditions (A, V, and A+V). Each DVD in the audio plus visual condition included 30 stimuli with congruent auditory and visual components. The other 30 stimuli were discrepant, used to elicit McGurk-like responses. Participants were asked to listen/watch each DVD and to verbally respond to what they perceived. No feedback was provided.

The pre-test was followed by five AV training sessions, each including one DVD from each talker. Trial-by-trial feedback was provided to the participants. For congruent stimuli, 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 the participant with a head nod. For discrepant stimuli, auditory component feedback was given to the listeners. For these stimuli, there is no "correct" response. The choice to provide the auditory component as feedback was made to determine whether listeners would become more reliant on the auditory portion of the stimulus over the course of training.

A mid-test, similar to the pre-test, was administered following the first five training sessions. No feedback was provided. Five more auditory plus visual training sessions, similar to the first five sessions, were administered after the mid-test. Finally, a post-test was conducted, without feedback. Testing and training took approximately 8-10 hours for each participant, and was broken up into two-to-three hour sessions. Participants were encouraged to take breaks to reduce fatigue.

Results and Discussion

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

Percent Correct Performance

Figure 1 shows the overall percent correct performance for congruent stimuli, 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. A slight increase in auditory scores was observed from pre- to post-test, as well as significant improvement in audio+visual scores. However, visual scores decreased from pre- to post-test. These results suggest that training the listeners in the audio+visual condition does produce an improvement in A+V performance. These results also imply that integration is a process independent of the auditory and visual conditions, given that the individual modality scores only changed slightly across tests. A two-factor, repeated-measures ANOVA showed no significant main effect of test (pre, mid, post), ($F(2,8)=1.144$,ns). However, a significant main effect of modality was observed, ($F(2,8)=66.128$, $p<.001$), as well as a significant interaction effect between test and modality, ($F(4,16)=5.511$, $p=.006$).

Figure 2 shows auditory, visual, and audio+visual pre-test responses averaged across listeners, for each talker. Four out of the five talkers were most intelligible in the audio+visual condition, followed by the visual condition, then the auditory condition. Interestingly, talker LG was perceived better in the auditory condition than in the visual. Talker LG also had considerably higher audio+visual intelligibility than the other talkers.

Figure 3 shows auditory, visual, and audio+visual mid-test responses, averaged across listeners, for each talker. In this case, three out of the five talkers were perceived better in the auditory condition rather than the visual condition. Also, intelligibility in the audio+visual modality increased in talkers DA, EA, JK, and KS.

Figure 4 shows auditory, visual, and audio+visual post-test responses, averaged across listeners, for each talker. Again, intelligibility in the audio+visual modality slightly increased for all talkers. At the post-test, all talkers were perceived better in the auditory condition than in the visual. This slight improvement in the auditory intelligibility for all talkers suggests that some learning in the auditory condition occurred from pre-test to post-test. However, the improvement was not statistically significant.

Figure 5 shows the amount of audio+visual integration, where integration is defined as the difference between audio+visual performance and the better single modality, auditory or visual, averaged across listeners, for each talker. A paired samples t-test, ($t(4)=10.56$, $p<.001$), showed a significant change in integration from pre-test to post-test for all talkers, suggesting that listeners do benefit from training. Improvement did not vary across talkers.

Figure 6 shows the amount of audio+visual integration for individual listeners. Four of the five listeners showed improvement in integration; however, the amount of improvement varied considerably. A possible explanation for this variability could be the difference in length of the time period that elapsed between training sessions for different listeners. Since participants were tested and trained at their convenience, some of these

sessions occurred over a shorter time span than others. Also, some participants seemed more motivated than others to challenge themselves and improve their scores by the end of the study.

Confusion matrices were calculated in order to see if certain stimuli benefited more from training than others. Table 1 shows confusion matrices for audio+visual performance, in the pre-test and post-test conditions, averaged across listeners and talkers. The confusion matrices indicate how stimuli were perceived before and after training. Results show an increase in the percent correct for 7 out of the 8 stimuli, and a decrease for one of the stimuli. The largest percent correct increases from pre-test to post-test were from the stimuli gat and tat. A slight decrease was seen for the syllable bat. Overall, when listeners did pick the incorrect stimuli, a large percent of the incorrect responses were from the appropriate viseme category. For example, the syllable bat was commonly mistaken for the syllable mat. This suggests a reliance on the visual information.

Integration of Discrepant Stimuli

In addition to congruent stimuli, listeners were also presented with discrepant stimuli, where the auditory stimulus differs from the visual stimulus. There is no "correct" response for these stimuli. These responses were 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.

Figure 6 shows overall percent discrepant responses for all tests, averaged across talkers and listeners. In the pre, mid, and post-tests listeners relied heavily on the visual modality. Also, listeners' responses did not significantly change from pre- to post-test. Interestingly, although participants were given the auditory component as feedback during training, this feedback did not increase their reliance on the auditory modality. These results indicate that training on a specific modality is limited to that modality and does not generalize.

Figure 7 further analyzes the "other" responses from Figure 6, and shows percent McGurk-type integration for discrepant responses, averaged across talkers and listeners. Listeners showed the highest percentage of fusion responses followed by neither responses. No combination responses were recorded. Due to the type of feedback provided during training sessions, no improvement in McGurk-type integration across training sessions was expected. Future studies should investigate training that provides McGurk-type integration feedback for the discrepant trials. Providing listeners with this type of feedback may produce an increase in McGurk-type responses for integration training. Perhaps because of the feedback structure of the present study, increases in integration efficiency for the congruent stimuli did not impact responding for the discrepant stimuli. This result may suggest that even integration training in one situation may not increase integration in other situations.

Chapter 4: Summary and Conclusion

Results of testing indicated that training in the audiovisual condition does result in a significant improvement in audiovisual integration skills. Four out of the five participants showed benefits from training, some considerably more than others. This suggests that while integration is a process that can be trained, personal factors or training schedules may play a major role in the amount of benefit that can be gained. While audio+visual scores increased, there were no improvements seen in the auditory-only or visual-only condition. These results support Grant and Seitz's (1998) argument that integration is a skill that is separate from processing in either individual modality.

Results also suggest that training on a specific modality is limited to that modality and does not generalize. Future studies should investigate training that provides McGurk-type integration feedback for the discrepant trials. Providing listeners with fusion responses as the "correct" answer may train listeners to integrate better, rather than favoring their better modality. This may also show an increase in McGurk-type responses for integration training. In addition, more specific investigation of the generalizability of integration training is needed.

Understanding the type of stimuli that best improves integration skills is essential in the design of aural rehabilitation programs in training hearing impaired persons to make use of any residual hearing. While it is possible to train integration, some listeners may still rely on their better single modality in some circumstances, or may encounter some situations involving the auditory-only or visual-only modality. Therefore, the most effective aural rehabilitation program would be one that trains auditory-only, visual-only,

and audio+visual conditions. To examine the independence of integration processing from a physiological perspective, imaging techniques, such as fMRI, might be employed. In any case, a listener with an aural rehabilitation program that is specialized to their needs and skills will be one that shows the most improvement.

Chapter 5: References

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

Exner, M. (2008). *Training Effects in Audio-Visual Integration of Sine Wave Speech.* Senior Honors Thesis, The Ohio State University.

Feleppelle, N. (2008). The role of the auditory signal in auditory-visual integration. Capstone Graduate Project, The Ohio State University.

Gariety, M. (2009). Effects of training on intelligibility and 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.

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

James, K. (2009). The effects of training on intelligibility of reduced information speech stimuli. Senior Honors Thesis, The Ohio State University.

Ladefoged, P. (2006). *A Course in Phonetics-Fifth Edition.* Boston : Wadsworth.

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

Munhall, K.G., Kroos, C., Jozan, C., & Vatikiotis-Bateson, E. (2004). Spatial frequency requirements for audiovisual speech perception. *Perception and Psychophysics, 66,* 574-583.

Remez, R.E., Rubin, P.E., Pisoni, D.B., & Carrell, T.D. (1981). Speech perception without traditional speech cues. *Science, 212 (4497),* 947-950.

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.

Tamosuinas, M. (2007). *Auditory-visual integration of sine-wave speech.* Senior Honors Thesis, The Ohio State University.

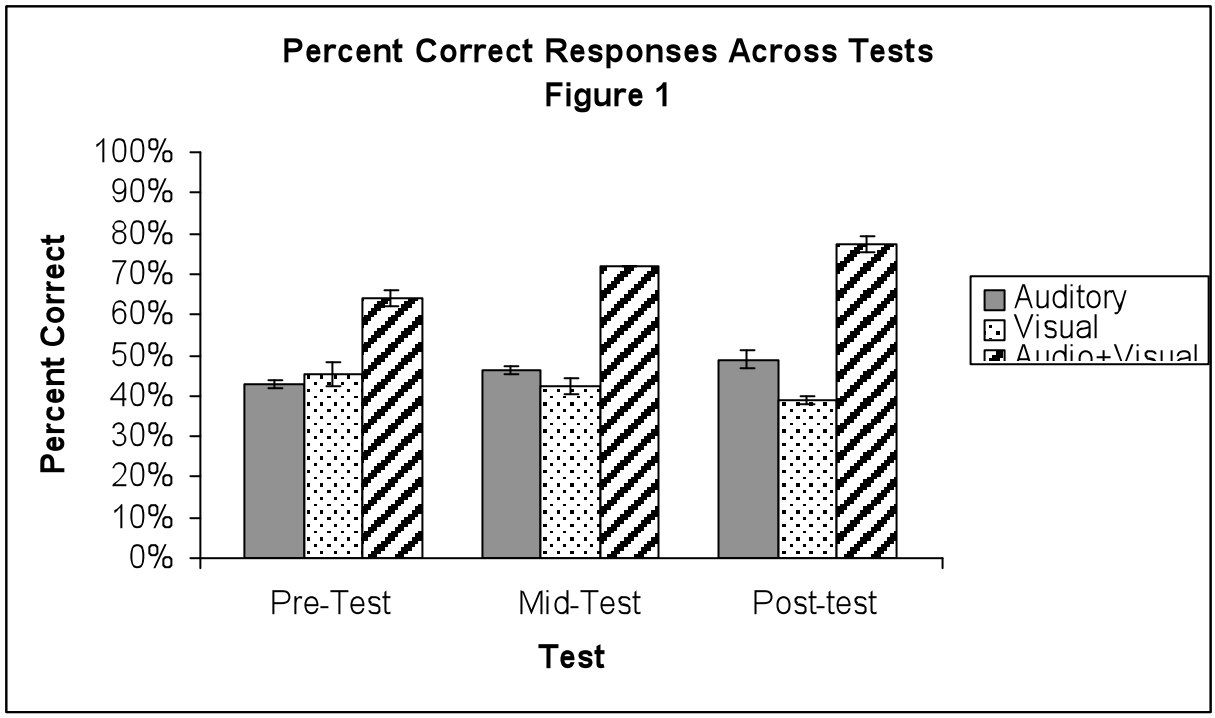
Table 1: Pre-test Audio+visual

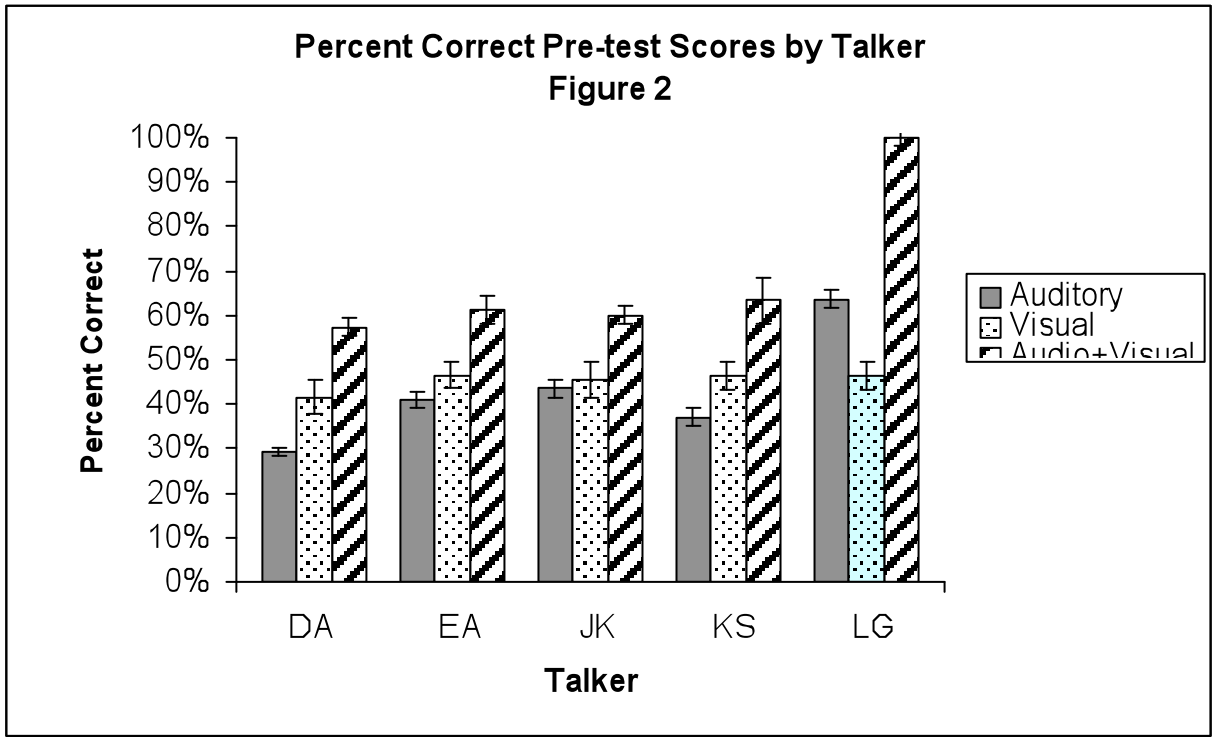
Stimulus \ Response	bat	pat	mat	gat	cat	sat	zat	tat
bat	**72%**	18%	10%	0%	0%	0%	0%	0%
pat	22%	**71%**	2%	0%	1%	0%	0%	4%
mat	19%	12%	**69%**	0%	0%	0%	0%	0%
gat	2%	0%	0%	**65%**	24%	0%	6%	3%
cat	0%	9%	0%	8%	**70%**	0%	0%	13%
sat	0%	0%	0%	2%	6%	**76%**	13%	3%
zat	2%	2%	0%	9%	5%	14%	**58%**	10%
tat	0%	4%	0%	5%	41%	6%	3%	**41%**

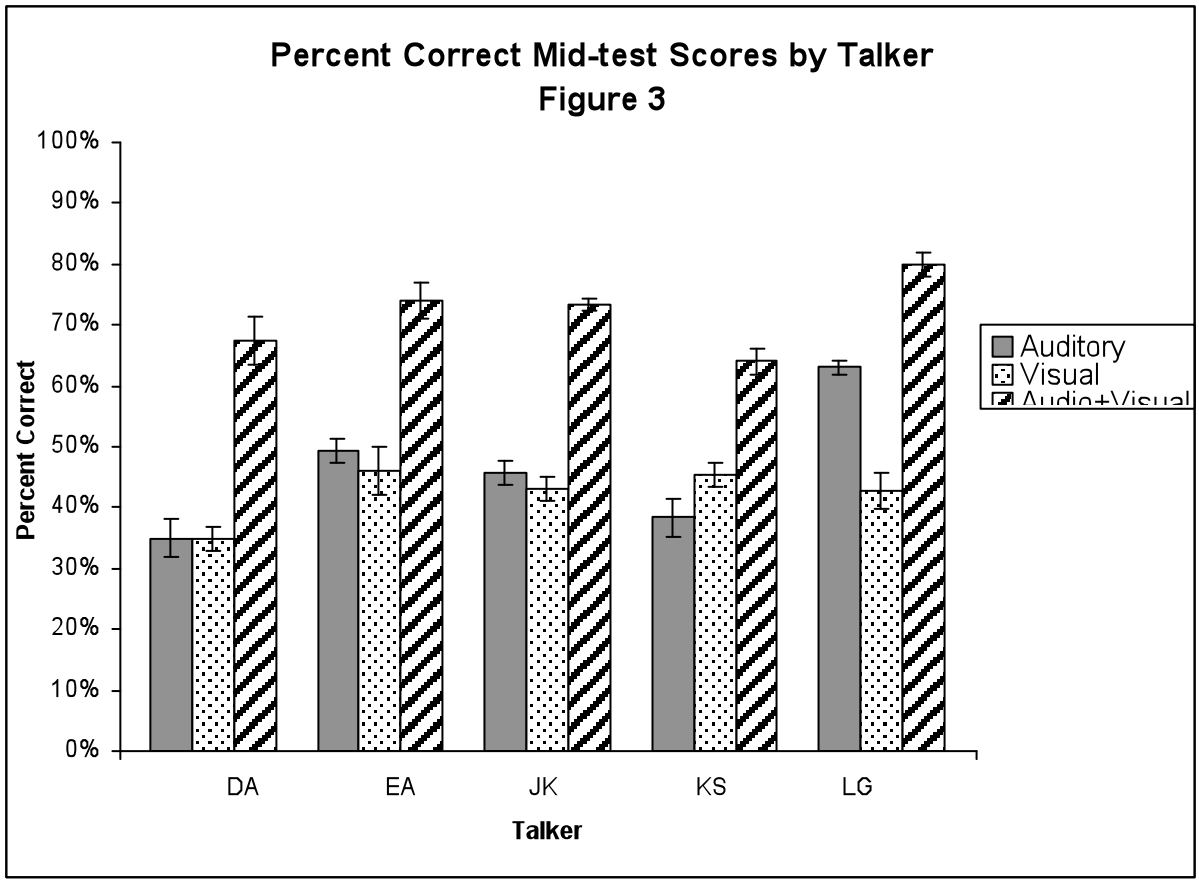
Post-test Audio+visual

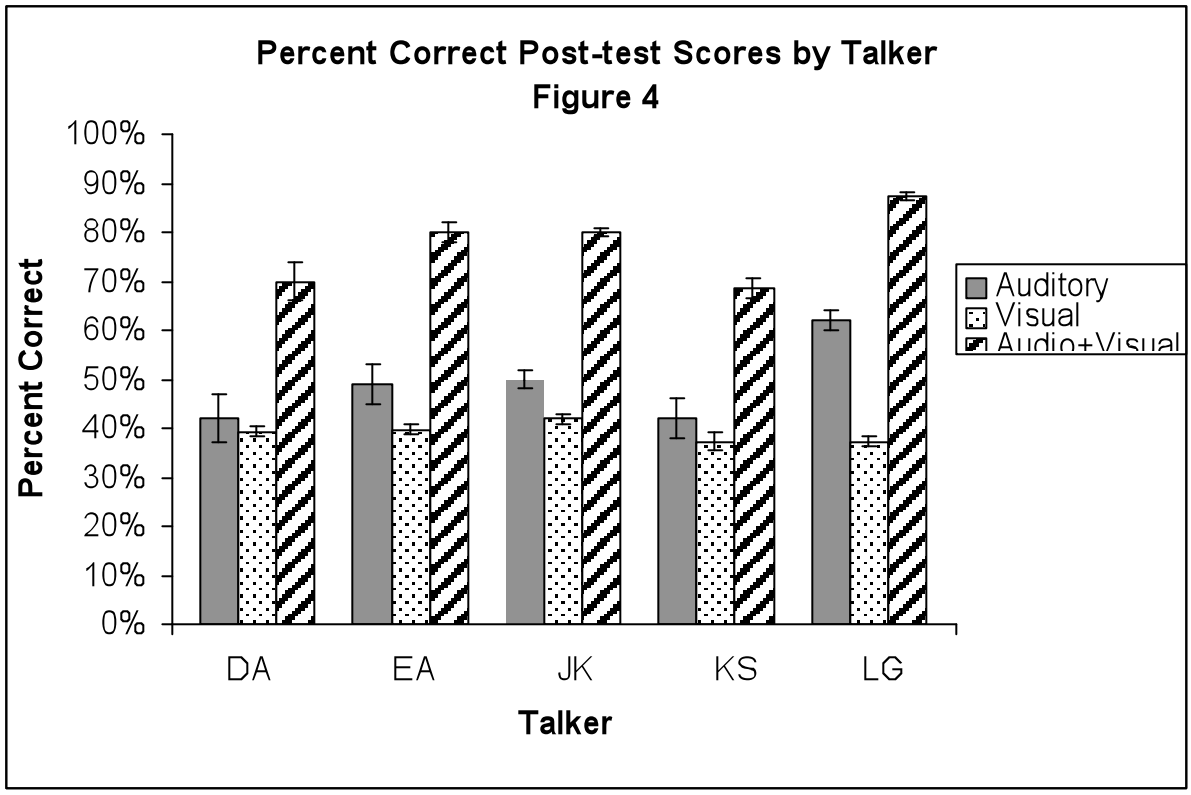
Stimulus \ Response	Bat	pat	Mat	Gat	Cat	Sat	Zat	Tat
Bat	**67%**	6%	26%	0%	0%	0%	0%	1%
Pat	14%	**76%**	8%	0%	1%	0%	0%	1%
Mat	5%	1%	**94%**	0%	0%	0%	0%	0%
Gat	0%	0%	0%	**76%**	20%	0%	2%	2%
Cat	1%	3%	0%	9%	**76%**	0%	1%	10%
Sat	0%	0%	0%	0%	1%	**89%**	9%	1%
Zat	0%	1%	0%	6%	0%	14%	**73%**	6%
Tat	0%	1%	0%	2%	19%	8%	1%	**69%**

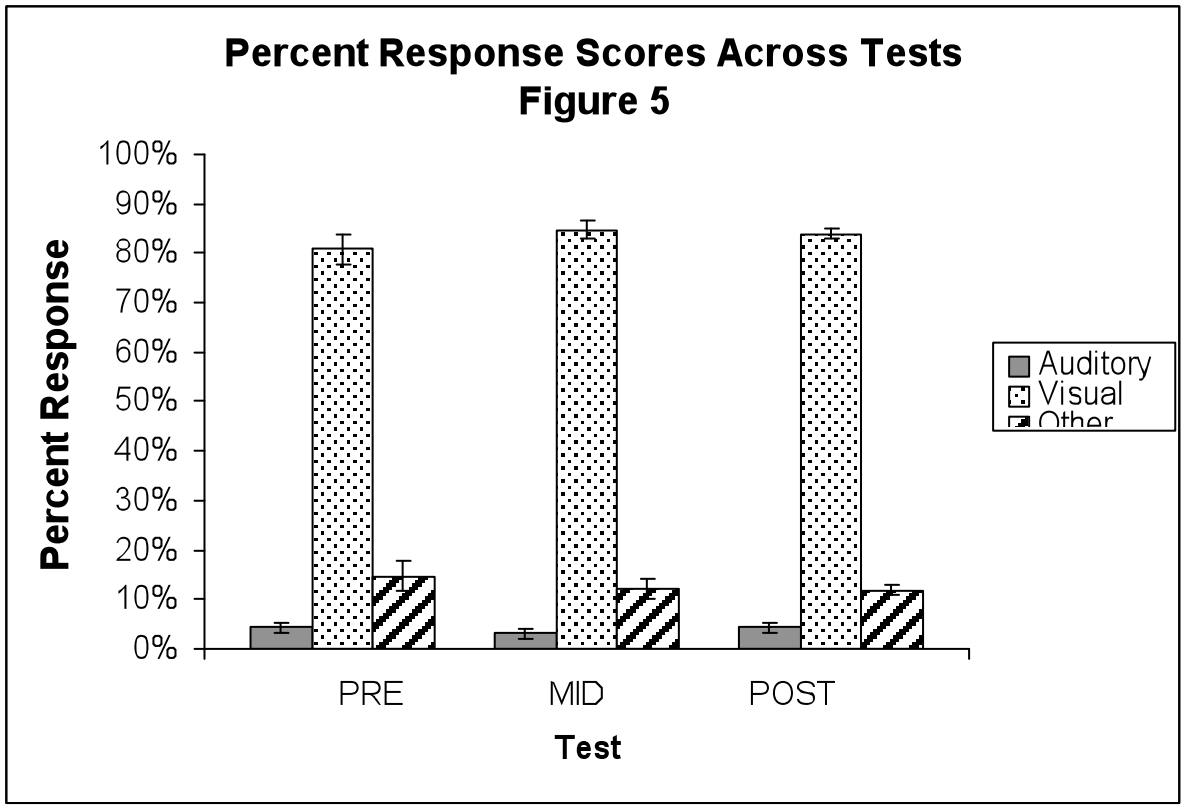
List of Figures

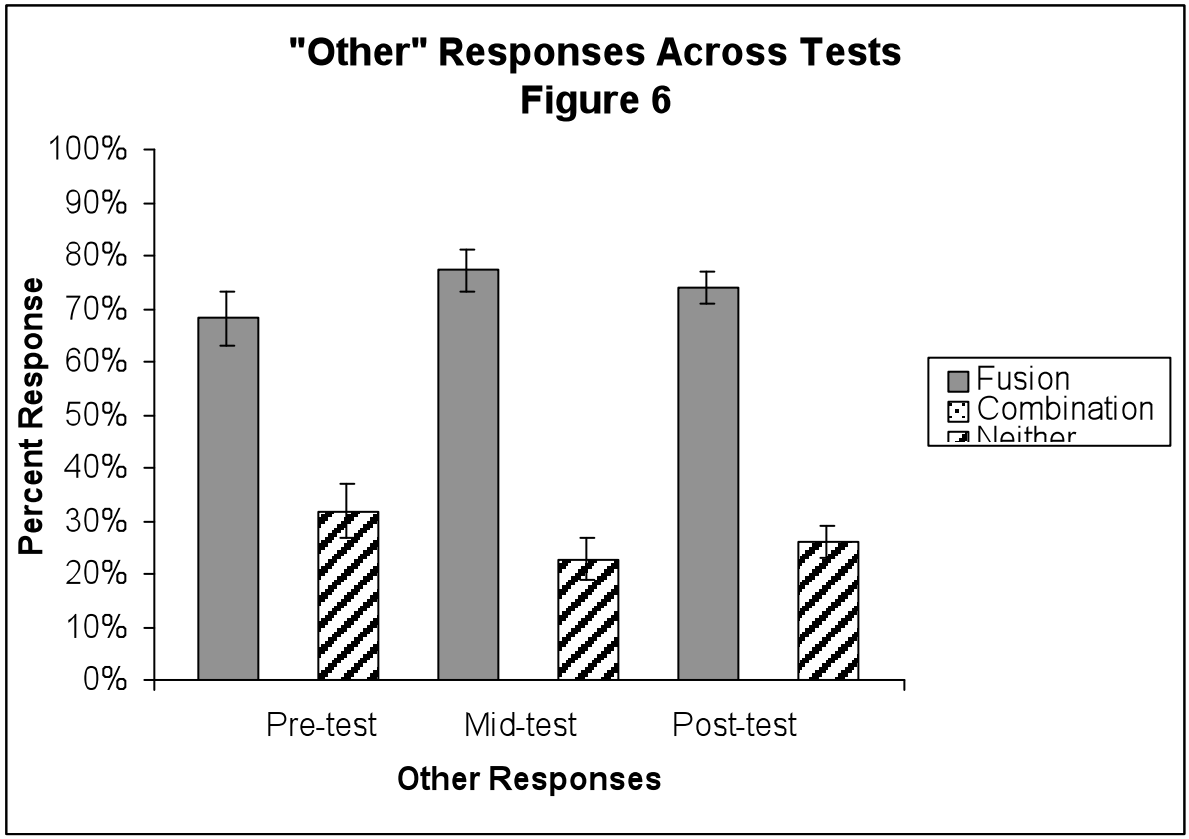

Percent Correct Responses Across Tests
Figure 1
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Percent Correct
Pre-Test
Mid-Test
Post-test
Test
Auditory
Visual
Audio+Visual

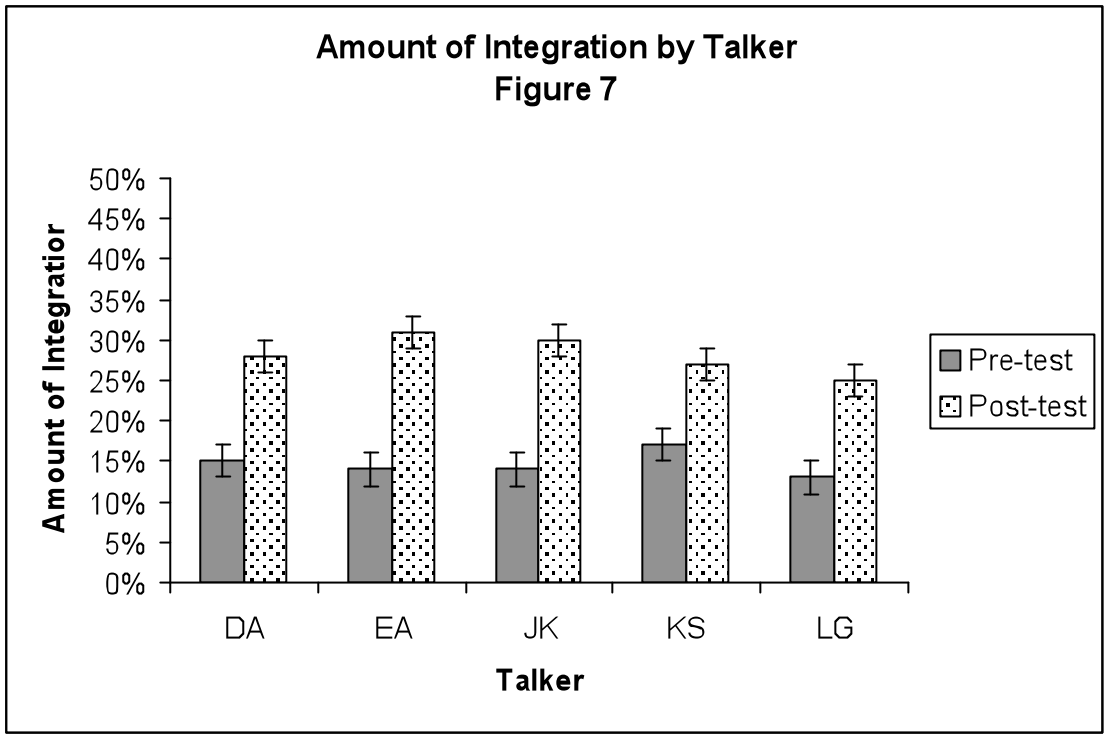

Percent Correct Pre-test Scores by Talker
Figure 2
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Percent Correct
DA
EA
JK
KS
LG
Talker
Auditory
Visual
Audio+Visual

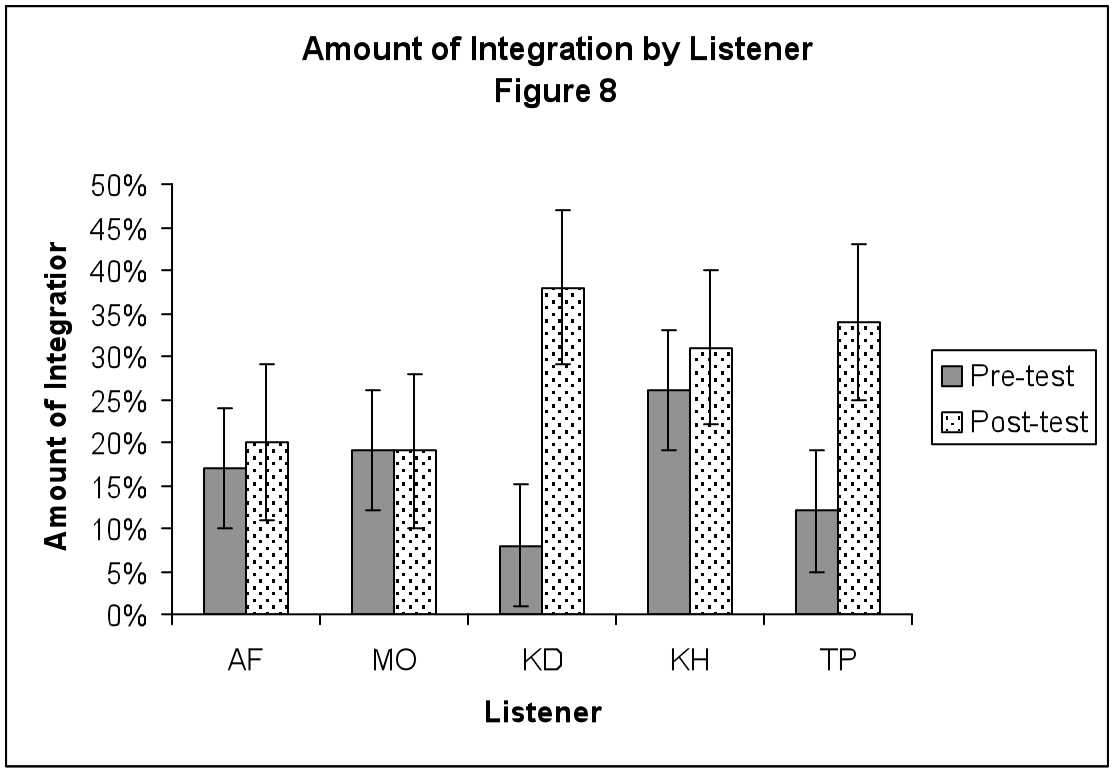

Percent Correct Mid-test Scores by Talker
Figure 3
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Percent Correct
DA
EA
JK
KS
LG
Talker
Auditory
Visual
Audio+Visual


Percent Correct Post-test Scores by Talker
Figure 4
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Percent Correct
DA
EA
JK
KS
LG
Talker
Auditory
Visual
Audio+Visual


Percent Response Scores Across Tests
Figure 5
Percent Response
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
PRE
MID
POST
Test
Auditory
Visual
Other


"Other" Responses Across Tests
Figure 6
Percent Response
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Pre-test
Mid-test
Post-test
Other Responses
Fusion
Combination
Neither


Amount of Integration by Talker
Figure 7
Amount of Integration
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
DA
EA
JK
KS
LG
Talker
Pre-test
Post-test


Amount of Integration by Listener
Figure 8
Amount of Integration
50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
AF
MO
KD
KH
TP
Listener
Pre-test
Post-test