# Visual Reliance and Visual Advantage

Visual input improves speech comprehension when auditory signals are degraded due to background noise (Sumby & Pollack, 1954) or hearing impairments (Walden, Prosek, & Worthington, 1975). Listeners strategically put greater focus on visual information to augment impoverished auditory information. This advantage of visual reliance is routinely utilized at clinical settings for adults with neurogenic communication disorders. Individuals experiencing auditory comprehension deficits are frequently encouraged by clinicians to look at speakers' faces during daily conversations. Despite the presumed advantage of visual information and the routine recommendation on the use of visual cues, there is a lack of research regarding the way individuals with brain lesions utilize visual information.

To take advantage of visual cues, listeners should be able to interpret facial movements into linguistically meaningful codes. Individuals who cannot efficiently process acoustic signals of speech may lack the ability of translating visual information into linguistic symbols (Schmid & Ziegler, 2006). In such cases, combining information from the two faulty channels may not enhance speech understanding. Even if an individual can recognize and interpret visual information, he/she may have a lesion in the cross-modal integration area (Miller & D'Esposito, 2005; Molholm et al., 2006), which will not allow the person to benefit from visual input.

This paper presents preliminary data from a study that examines: (1) whether individuals with stroke rely on visual input when auditory information is ambiguous, (2) whether an enforced reliance on visual cues can improve speech understanding, and (3) whether a participant's ability to decipher visual information can be accounted for by his/her cognitive-linguistic characteristics.

## Method

### *Participants*

Seven individuals who had experienced stroke participated in the study. The current paper presents preliminary data from four of the participants. Table 1 presents demographic profiles and aphasia quotients (Kertesz, 2007) of the four individuals. All participants demonstrated visual and auditory acuity appropriate for the study procedures.

### *Stimuli*

The participants' visual reliance and auditory-visual processing skills were measured using the McGurk paradigm (McGurk & MacDonald, 1976). The experimental stimuli included three non-word syllables (*ba*, *da*, and *ga*) and six monosyllabic words (*beer*, *deer*, *gear*, *bunk*, *dunk*, and *gunk*). These real word stimuli were included to observe a potential lexical effect on participants' performance. The nine stimuli were recorded using a camcorder (Canon FS 100). To capture high-quality speech signals, an external microphone (Shure) was connected through a preamplifier (PreSonus TubePRE) to the camcorder for sound recording. The stimuli were spoken by a graduate research assistant, who was a female native speaker of English. Her face and shoulders were visible in video frames. Each video clip started and ended with a 500 ms segment, during which the speaker's face was static and her mouth closed.

The stimuli were presented in four conditions: auditory-only (A-only), visual-only (V-only), auditory-visual (AV), and Enforced auditory-visual. The stimuli for the A-only condition were generated by extracting sound signals from the video clips using Microsoft Windows

Movie Maker. This method ensured that the sound quality of the A-only stimuli was the same as that of the AV stimuli. The V-only condition presented the AV stimuli with the sound turned off.

The AV condition presented congruent (e.g., video *ga* – sound [ga]) and incongruent (e.g., video *ga* – sound [ba]) video clips. Incongruent stimuli were generated using Sound Forge 9 (Sony) by visually matching the onset of initial consonants in two waveforms (e.g., [g] and [b]) and then replacing the original sound segment (e.g., [g]) with an incongruent sound segment (e.g., [b]). In total, each AV condition included five stimuli (e.g., congruent *ba*, congruent *da*, congruent *ga*, incongruent *ba* – [ga], and incongruent *ga* – [ba]).

The Enforced AV condition presented one odd stimulus in addition to the three congruent and two incongruent stimuli. The odd stimulus showed the speaker who was mouthing the word, *ice cream,* while the sound segment [ba] was being played.

*Procedure*

The experimental tasks were presented using Alvin software (Hillenbrand & Gayvert, 2005). The tasks were grouped into blocks by stimulus type (*ba-da-ga, beer-deer-gear, & bunk-dunk-gunk)* and conditions (A-only, V-only, AV, Enforced AV). Thus, a participant completed 12 blocks of tasks (= 3 stimulus types x 4 conditions). In each block, individual stimuli were presented ten times in a randomized order.

In the A-only, V-only, and AV conditions, participants were given three choices for response. The non-word syllables (i.e., *ba, da,* and *ga)* were visually presented in written form. The real words (i.e., *beer, deer,* and *gear; bunk, dunk,* and *gunk*) were presented in written form along with corresponding color pictures. Participants were asked to point to their choice for each stimulus and/or verbally repeat it.

In the Enforced AV condition, a color picture of ice cream was presented along with the other three choices. Participants were required to spot all odd items as well as identify the other stimuli. Thus, participants were enforced to focus on the speaker's face.

*Analysis*

Responses in the A-only and V-only conditions were coded as correct or incorrect. Responses to congruent items in the AV and Enforced AV conditions were coded as correct or incorrect. Responses to incongruent items (e.g., video *ga* – sound [ba]) were coded as fusion (e.g., [da]), visual bias (e.g., [ga]) or auditory bias (e.g., [ba]). The percentage of fusion and visual bias quantified the visual reliance of the participants.

## Results and Discussion

Table 2 summarizes accuracy and visual reliance data. Visual reliance is greater in the Enforced AV condition than in the AV condition for three of the four participants. However, not all participants' accuracy scores are higher in the Enforced AV condition than in the AV condition. With the enforced visual attention, one participant (P3) improves her AV accuracy, whereas two participants (P4 & P5) shows decline. It should be noted that P3 has a higher score than P4 and P5 in the V-only condition. These data suggest that increased attention to visual information does not necessarily improve speech understanding. Individuals with poor visual processing skills may experience a detrimental effect of enforced visual attention.

A Pearson correlation analysis on the data from individual blocks reveals that the accuracy and the visual reliance in the AV condition are strongly associated in a negative direction ($r^2$ = -828, $p < .01$). That is, the higher the accuracy is, the less the visual reliance is.

Two potential explanations for this finding are: (1) When auditory signals are clear, listeners do not need to rely on visual cues; and (2) Attention to visual information interferes with the cognitive processing of acoustic signals. These hypotheses should be further tested in subsequent experimental studies.

Additional analyses are underway to examine: the data from the other three participants; results from all seven participants on published tests, *Western Aphasia Battery – Revised* (Kertesz, 2007), the *Cognitive Linguistic Quick Test* (Helm-Estabrooks, 2001), and the *Apraxia Battery for Adults – Second Edition* (Dabul, 2000); and the seven participants' auditory-visual processing skills on real words, short sentences, and long sentences. These additional data, along with the current findings, will explicate the association between cognitive-linguistic skills, visual reliance, and visual advantage. The results may suggest that at least for some individuals with auditory comprehension deficits, "Watch me say it" may not be an instruction that is as effective as "Listen to me carefully."

References

Dabul, B. (2000). *Apraxia Battery of Adults – Second Edition*. Austin, TX: Pro-Ed.

Helm-Estabrooks, N. (2001). *Cognitive Linguistic Quick Test*. San Antonio, TX: Harcourt Assessment Company.

Hillenbrand, J.M. & Gayvert, R.T. (2005). Open source software for experiment design and control. *Journal of Speech, Language, and Hearing Research*, 48, 45-60.

Kertesz, A. (2007). *Western Aphasia Battery – Revised*. San Antonio, TX: Harcourt Assessment Company.

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

Miller, L.M., & D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *Journal of Neuroscience*, 25 (25), 5884-5893.

Molholm, S., Sehatpour, P., Mahta, A.D., Shpaner, M., Gomez-Ramirez, M., Ortigue, S., et al. (2006). Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. *Journal of Neurophysiology*, 96, 721-729.

Schmid, G., & Ziegler, W. (2006). Audio-visual matching of speech and non-speech oral gestures in patients with aphasia and apraxia of speech. *Neuropsychologia*, 44, 546-555.

Sumby, W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26 (2), 212-215. 11-16.

Walden, B.E., Prosek, R.A., & Worthington, D.W. (1975). Auditory and audiovisual feature transmission in hearing-impaired adults. *Journal of Speech and Hearing Research*, 18 (2), 272-280.

Table 1
*Demographic Characteristics and Aphasia Quotients of the Study Participants*

| Participant Code | Etiology | Gender | Age | Time Postonset (yrs; mos) | Years of Education | Aphasia Quotient |
|---|---|---|---|---|---|---|
| P3 | L CVA | F | 79 | 5;2 | 18 | 83.4 |
| P4 | L CVA | F | 57 | 12;0 | 14 | 93.2 |
| P5 | L CVA | M | 72 | 9;3 | 21 | 73.7 |
| P6 | R CVA | M | 65 | 0;11 | 14 | 94.2 |

Table 2
*Accuracy and Visual Reliance Data of Four Participants*

| Participant Code | Accuracy (%) | | | | Visual Reliance (%) | |
|---|---|---|---|---|---|---|
| | A-only | V-only | AV | Enforced AV | AV | Enforced AV |
| P3 | 98.9 | 73.3 | 95.6 | 98.9 | 20.0 | 60.0 |
| P4 | 100.0 | 52.2 | 100.0 | 86.7 | 1.7 | 0.0 |
| P5 | 88.9 | 44.4 | 93.3 | 87.8 | 28.3 | 30.0 |
| P6 | 97.8 | 86.7 | 100.0 | 100.0 | 15.0 | 26.7 |