Audiovisual Integration of Point Light Displays of Speech by Deaf Adults following Cochlear Implantation

Tonya R. Bergeson†, David B. Pisoni†, Lorin Lachs‡ and Lindsey Reese†

† Indiana University School of Medicine, Department of Otolaryngology, 699 West Street – RR044, Indianapolis, Indiana, USA 46202

‡ California State University-Fresno, Department of Psychology, 5310 North Campus Drive M/S PH11, Fresno, California, USA 93740-8019

E-mail: tbergeso@iupui.edu, pisoni@indiana.edu, llachs@csufresno.edu

ABSTRACT

Normal-hearing (NH) adults display audiovisual enhancement when degraded auditory input (e.g., words, sentences) is paired with point-light displays of speech, which isolate the kinematic properties of a speaker's face [10]. Do deaf adults who use cochlear implants (CIs) benefit in the same way? In the present study, we investigated audiovisual (AV) word recognition using point-light displays (PLDs) of speech in a small group of postlingually deaf adults with CIs and a group of NH adults. Participants were asked to repeat aloud what they thought the talker said under three conditions: Auditory-alone, Visual-alone, and Audiovisual. Both groups displayed evidence of AV enhancement with PLDs. These results suggest that NH and CI adults were sensitive to the kinematic properties in speech represented in the PLDs, and they were able to use kinematics to improve their word recognition performance even with highly degraded visual displays of speech.

1. INTRODUCTION

Visual information about speech articulation obtained from lipreading has been shown to improve speech perception in adults with normal hearing [1, 2], hearing loss [3], and deaf adults with cochlear implants (CIs) [4, 5, 6]. In fact, many audiologists and speech and hearing scientists have assumed that the primary modality of speech perception is vision for hearing-impaired people [1, 7, 8].

In this connection, it has been proposed recently that hearing-impaired adults who are highly successful lipreaders exhibit larger audiovisual benefit than those who are poor lipreaders [4]. What are the cues that hearing-impaired adults attend to while lipreading? It is possible that good lipreaders are more sensitive to the changes in time, or kinematics, common to both auditory and visual speech patterns. One method of assessing sensitivity to time-varying visible speech information is to use point-light displays (PLDs). This method involves placing small point-lights on target locations of a darkened actor's face, videotaping the actor articulating a list of words, and then playing back the videotapes to individuals so that only the movement of the lights can be seen. When presented statically, point-light displays cannot be recognized as a human face. However, once the point-lights begin to move and there is change over time, observers are able to recognize the displays as a human face articulating words. Thus, PLDs can be used to isolate the kinematic properties of the visual speech signal [9, 10].

Several studies have shown that normal-hearing (NH) adults display AV enhancement when degraded auditory input (e.g., words, sentences) is paired with PLDs of speech [10, 11, 12]. Do deaf adults who use CIs benefit in the same way? In the present study, we investigated audiovisual (AV) word recognition using point-light displays of speech in a small group of postlingually deaf adults with CIs.

2. METHOD

2.1 Participants

Five cochlear implant (CI) participants were deafened after the age of five, had used their CIs for at least one year, and were between the ages of 36 and 74 years ($\underline{M}$ = 56.2 years). Six normal-hearing (NH) participants were also included as a control group, ranging from 22 to 24 years of age ($\underline{M}$ = 23.0 years).

	Age at Test (years)	Age at Implantation (years)	Duration of Implant Use (years)
CI 18	39	30	8
CI 50	74	73	1
CI 80	62	53	9
CI 94	36	35	1
CI 95	70	65	4

Table 1: Characteristics of CI participants.

ISBN 1-876346-48-5

2.2 Stimuli and Procedure

The point-light displays of speech were constructed by darkening the face of the talker and by placing 10 dots symmetrically around the lips and mouth of the talker, 2 dots on the chin, 8 dots along the jaw line, 2 dots on the cheeks, 1 dot on the tip of the nose, 2 dots on the upper teeth, 2 dots on the lower teeth, and 1 dot on the talker's tongue (28 points total) (see Figure 1). The talker was videotaped while reading isolated English words under an infrared light so only the reflective disks surrounding the talker's articulators could be seen. The talkers in the full-face and point-light displays were two different women. The full-face and point-light conditions each contained a different set of 96 monosyllabic English words, equally divided into words with high and low visual intelligibility.

Both groups of participants were instructed to repeat aloud what they thought the talker said under three presentation conditions: Auditory-alone (A-alone), in which the words were presented via a sound speaker while the computer screen remained blank, Visual-alone (V-alone), in which visual displays of the talker articulating words were presented on the computer screen while the speaker was off, and Auditory-visual (AV), in which the words were presented via the sound speaker and the computer screen. A constant background noise (55 dB SPL) was present in the testing room. We presented the auditory stimuli at 75 dB SPL for CI users, but at 60 dB SPL for NH listeners so that gains due to visual input could be observed.

There were three phases of the present study: First, both groups of participants were initially given a practice session using a full-face display. Following the practice session, participants were then given the word recognition tests first with the full-face display, and finally with the point-light display.

3. RESULTS AND DISCUSSION

The data were scored by the number of whole words, phonemes, and visemes correctly identified. A viseme is a visual category of speech consisting of speech sounds that look similar when articulated [13]. For example, one viseme group might contain "va" and "fa" while another viseme group comprises "ma," "ba," and "pa".

Figures 2 and 3 show the proportion correct for word, phoneme, and viseme recognition in CI and NH participants across the two visual displays (full-face and point-light) and three presentation conditions (A-alone, V-alone, and AV). When CI participants' responses were scored by words correctly identified, we found statistically significant main effects of visual display ($F(1, 4) = 24.79$, $p < .01$) and presentation format ($F(1, 4) = 69.60$, $p = .001$). The interaction between visual display and presentation format was not statistically significant. As shown in Figure 2, CI participants' performance was better in the full-face condition than in the point-light condition. Performance was best in the AV presentation condition, followed by the A-alone presentation condition, and then the V-alone presentation condition.

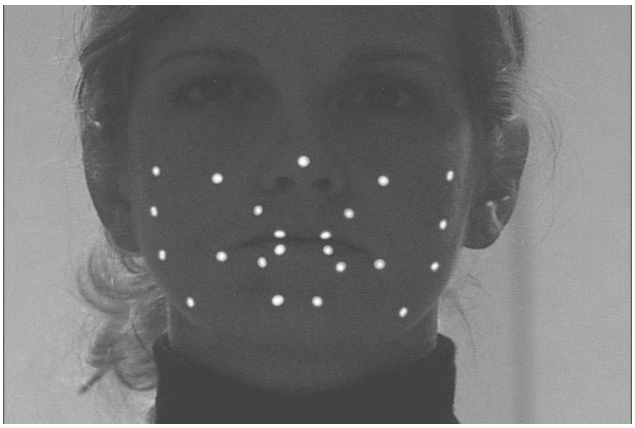

Figure 1: Dot configuration for point-light display. Five dots are not visible due to occlusion by the lips.

When NH participants' responses were scored by words correctly identified, we found a statistically significant main effect of presentation mode ($F(1, 5) = 9.76$, $p < .05$), but no significant main effect of visual display, and no significant interaction between visual display and presentation format. As shown in Figure 3, performance for NH participants was best in the AV presentation condition, followed by the A-alone presentation condition, and then the V-alone presentation condition, similar to the pattern of performance for CI participants. Performance was not better in the full-face condition compared to the point-light display condition.

When CI participants' responses were scored by the proportion of phonemes correctly identified, we found a statistically significant main effect of visual display ($F(1, 4) = 88.85$, $p = .001$) and presentation format ($F(1, 4) = 54.62$, $p < .01$). The interaction was not statistically significant. Again, as shown in Figure 2, CI participants' performance was better in the full-face condition than in the point-light condition. Although performance was best in the AV presentation condition, there was not a clear advantage for performance in the A-alone presentation condition compared to the V-alone presentation condition.

When NH participants' responses were scored by the proportion of phonemes correctly identified, we also found a statistically significant main effect of visual display ($F(1, 5) = 48.34$, $p = .001$) and a marginally significant main effect of presentation format ($F(1, 5) = 5.18$, $p = .072$). The interaction was not statistically significant. As shown in Figure 3, NH participants' performance was best in the AV presentation condition, followed by the A-alone presentation condition, and then the V-alone presentation condition. Performance was slightly better in the full-face condition than in the point-light display condition.

Finally, when CI participants' responses were scored by proportion of visemes correctly identified, we found

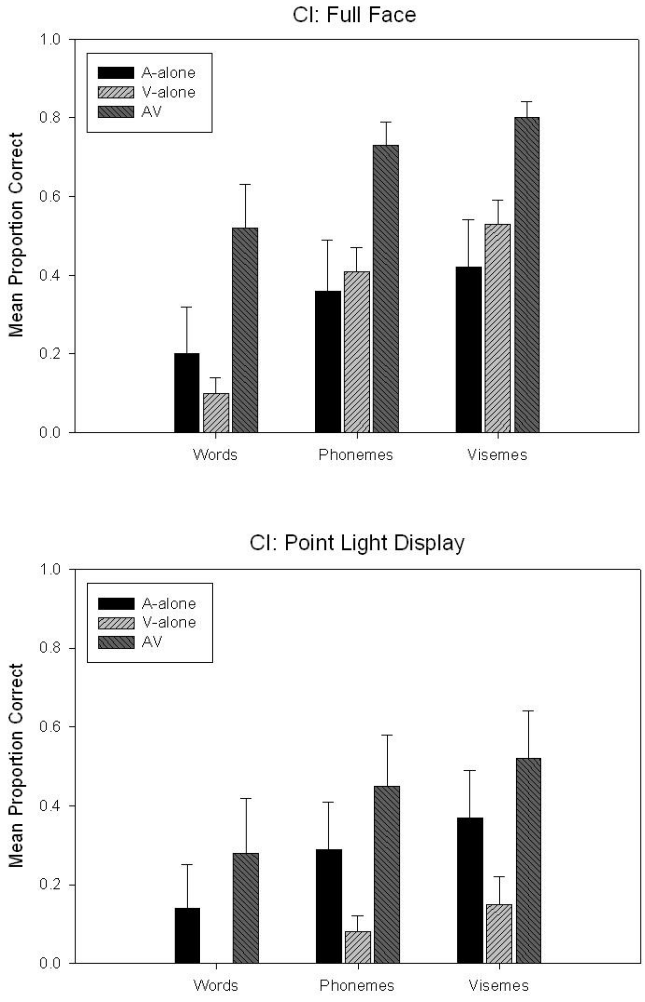

Figure 2: Word, phoneme, and viseme recognition performance for CI adults in A-alone, V-alone, and AV conditions for full-face versus point-light visual display.

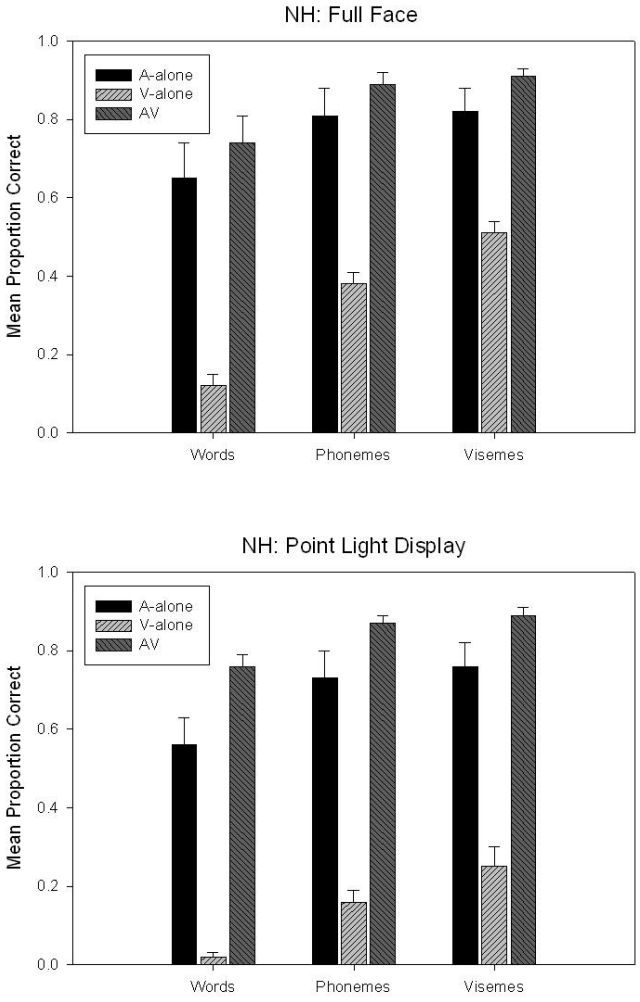

Figure 3: Word, phoneme, and viseme recognition performance for NH adults in A-alone, V-alone, and AV conditions for full-face versus point-light visual display.

statistically significant main effects of visual display ($F(1, 4) = 50.41$, $p < .01$) and presentation format ($F(1, 4) = 37.77$, $p < .01$), as well as a statistically significant interaction between visual display and presentation format ($F(1, 4) = 7.79$, $p < .05$). Once again, CI participants' performance was better in the full-face condition than in the point-light condition. Performance was best in the AV presentation condition. In the point-light display condition, performance was better in the A-alone presentation condition than in the V-alone presentation condition. However, in the full-face display condition, performance was better in the V-alone presentation condition than in the A-alone presentation condition.

When NH participants' responses were scored by proportion of visemes correctly identified, we found a statistically significant main effect of visual display ($F(1, 5) = 26.96$, $p < .01$) and a marginally significant main effect of presentation format ($F(1, 5) = 5.61$, $p = .064$). The interaction was not statistically significant. As shown in Figure 3, NH participants' performance was best in the AV presentation condition, followed by the A-alone presentation condition, and then the V-alone presentation condition. Performance was slightly better in the full-face condition compared to the point-light display condition.

Note that V-alone word, phoneme, and viseme recognition performance appears to be similar for NH and CI listeners across full-face and point-light displays. In fact, a two-tailed t-test revealed no statistically significant differences between NH and CI listeners in each condition. Vision has been assumed to be the primary modality of speech perception for hearing-impaired people [1, 7, 8]. The present results, however, show that the lipreading skills of this small group of postlingually deaf adult CI users are not better than the lipreading skills of NH adults.

It is important to mention that in the present study three of

ISBN 1-876346-48-5

the CI users had sudden-onset hearing impairment and two had progressive hearing impairment. The participants also varied greatly in terms of their age at onset of deafness, ranging from 11 years to 63 years. Recent studies have shown that deaf adults with CIs who had progressive hearing loss are better lipreaders than those with sudden hearing loss [14], and that adults with early-onset hearing loss are better lipreaders than those with late-onset hearing loss [5, 15, 16]. Thus, it is possible that with a larger sample size, we might find similar effects of hearing loss and age at onset of deafness on the V-alone performance using sentences.

4. CONCLUSIONS

The results showed that both groups of listeners displayed evidence of audiovisual enhancement. Overall, participants performed most accurately in the AV condition, followed by the A-alone condition, and then the V-alone condition. These results suggest that adult CI users, like NH adults, were sensitive to the kinematic properties in speech represented by the dynamic changes in the point-light displays, and they were able to use kinematics to improve their word recognition performance even with highly degraded visual displays of speech.

5. ACKNOWLEDGEMENTS

This work was supported by NIH-NIDCD Training Grant T32DC00012 to Indiana University and NIH-NIDCD Research Grant R01DC00111 to the Indiana University School of Medicine.

REFERENCES

[1] N. P. Erber, "Interaction of audition and vision in the recognition of oral speech stimuli", *Journal of Speech and Hearing Research*, **12**, pp. 423-425, 1969.

[2] W. H. Sumby, I. Pollack, "Visual contribution to speech intelligibility in noise", *Journal of the Acoustical Society of America*, **26**, pp. 212-215, 1954.

[3] N. P. Erber, "Auditory-visual perception of speech", *Journal of Speech and Hearing Disorders*, **40**, pp. 481-492, 1975.

[4] K. W. Grant, B. E. Walden, P. F. Seitz, "Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration", *Journal of the Acoustical Society of America*, **103**, pp. 2677-2690, 1998.

[5] A. R. Kaiser, K. I. Kirk, L. Lachs, D. B. Pisoni, "Talker and lexical effects on audiovisual word recognition by adults with cochlear implants", *Journal of Speech, Language, and Hearing Research*, in press.

[6] R. F. Tyler, A. J. Parkinson, G. G. Woodworth, M. W. Lowder, B. J. Gantz, "Performance over time of adult patients using the Ineraid or nucleus cochlear implant", *Journal of the Acoustical Society of America*, **102**, pp. 508-522, 1997.

[7] J.-P. Gagné, "Visual and audiovisual speech perception training", In J.-P. Gagné & N. Tye-Murray (Eds.), *Research in Audiological Rehabilitation: Current Trends and Future Directions (Monograph). Journal of the Academy of Rehabilitative Audiology*, **27**, pp. 133-159, 1994.

[8] R. C. Seewald, M. Ross, T. G. Giolas, A. Yonovitz, "Primary modality for speech perception in children with normal and impaired hearing", *Journal of Speech and Hearing Research*, **28**, pp. 36-46, 1985.

[9] G. P. Bingham, L. D. Rosenblum, R. C. Schmidt, "Dynamics and the orientation of kinematic forms in visual event recognition", *Journal of Experimental Psychology: Human Perception and Performance*, **21**, pp. 1473-1493, 1995.

[10] L. D. Rosenblum and H. M. Saldaña, "An audiovisual test of kinematic primitives for visual speech perception", *Journal of Experimental Psychology: Human Perception and Performance*, **22**, pp. 318-331, 1996.

[11] L. Lachs, D. B. Pisoni, K. I. Kirk, "Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report", *Ear & Hearing*, **22**, pp.236-251, 2001.

[12] L. D. Rosenblum, J. A. Johnson, H. M. Saldaña, "Point-light facial displays enhance comprehension of speech in noise", *Journal of speech and Hearing Research*, **39**, pp.1159-1170, 1996.

[13] B. E. Walden, R. A. Prosek, A. A. Montgomery, C. K. Scherr, & C. J. Jones, "Effects of training on the visual recognition of consonants", *Journal of Speech and Hearing Research*, **20**, pp. 130-145, 1977.

[14] T. R. Bergeson, D. B. Pisoni, L. Reese, K. I. Kirk, "Audiovisual speech perception in adult cochlear implant users: Effects of sudden vs. progressive hearing loss", Poster presented at the annual research meeting of the Association for Research in Otolaryngology, Daytona Beach, Florida, Feb, 2003.

[15] T. R. Bergeson, D. B. Pisoni, "Audiovisual speech perception in deaf adults and children following cochlear implantation", in G. Calvert, G. Spence, & B. E. Stein (Eds.) *Handbook of Multisensory Integration*, Cambridge, MA: MIT Press, in press.

[16] I. Tillberg, J. Rönnberg, I. Svärd, B. Ahlner, "Audio-visual speechreading in a group of hearing aid users: The effects of onset age, handicap age, and degree of hearing loss", *Scandinavian Audiology*, **25**, pp. 267-272, 1996.