RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 26 (2003-2004)
Indiana University

Multimodal Sentence Intelligibility and the Detection of Auditory-Visual Asynchrony in Speech and Nonspeech Signals: A First Report[1]

Brianna L. Conrey

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

---

[1] This work was supported by NIH NIDCD R01 research grant DC00111 and NIH NIDCD T32 training grant to Indiana University. The author was also supported by an NSF Graduate Research Fellowship. I would like to thank David Pisoni for his help and advice through all stages of this project. I would also like to thank Luis Hernandez for technical support and Sara Phillips for assistance with data collection.

Multimodal Sentence Intelligibility and the Detection of Auditory-Visual Asynchrony in Speech and Nonspeech Signals: A First Report

Abstract. The ability to perceive and understand visual-only speech and the benefit experienced from having both auditory and visual signals available during speech perception tasks varies widely in the normal-hearing population. At the present time, little is known about the underlying neural mechanisms responsible for this variability or the possible relationships between multisensory speech perception abilities and performance on other perceptual or cognitive tasks. Previous studies have hypothesized that lipreading ability and auditory-visual (AV) benefit measures might be positively correlated with the ability to detect asynchronies in AV signals. Good integrators might be more attuned to detailed temporal relationships between auditory and visual information. However, this hypothesis has not been explicitly tested in normal-hearing individuals. In the present investigation, 30 normal-hearing participants were given a modified clinical test of sentence intelligibility (the CUNY sentences) under auditory-only, visual-only, and auditory-visual (AV) presentation conditions. The same participants also performed an AV asynchrony detection task using both speech and nonspeech multimodal stimuli that varied over a range of temporal asynchronies. The results suggest a relationship between auditory-only, visual-only, and AV sentence intelligibility measures and the ability to detect AV asynchrony in speech and nonspeech signals. Implications for AV integration in speech perception are discussed.

Introduction

The ability to detect temporal asynchrony between auditory and visual signals varies considerably among individuals (e.g., Conrey & Pisoni, 2003, this volume; Grant & Seitz, 1998; McGrath & Summerfield, 1985). For example, McGrath and Summerfield reported that although some individuals are consistently able to detect auditory-visual (AV) asynchronies when the auditory signal precedes the visual signal by as little as 30 ms or when the visual signal precedes the auditory signal by as little as 70 ms, others require as much as 210 ms of auditory lead or 330 ms of visual lead.

Previous reports in the literature have suggested that individuals who are highly accurate at understanding visual-only speech (i.e., "good lipreaders") may also be better at detecting AV asynchronies than poor lipreaders, because good lipreaders might be more attuned to detailed temporal relationships between auditory and visual information (Grant & Seitz, 1998; McGrath & Summerfield, 1985; Pandey, Kunov, & Abel, 1986). In their 1985 study, McGrath and Summerfield presented participants with asynchronous AV sentences in which the auditory signal was only the fundamental frequency of the speaker's voice. They reported that good lipreaders displayed significantly decreased speech perception accuracy scores with auditory delays of 80 ms, but the performance of average and poor lipreaders did not decrease significantly until the auditory signal was delayed by 160 ms. This finding suggests that good lipreaders were less able to effectively integrate auditory and visual information when the two sources of information were asynchronous. In a follow-up experiment, McGrath and Summerfield used a three-interval forced-choice procedure and asked participants to identify the target synchronous stimulus from a pair of AV nonspeech approximations of CV syllables. They reported that higher lipreading scores had a moderate association with lower thresholds for detecting AV asynchrony, but this association did not reach statistical significance, perhaps because of the small number of subjects.

In another study, Pandey et al. (1986) also tested experienced and inexperienced lipreaders on the perception of AV sentences in which the auditory signal was delayed and mixed with multitalker babble. At a SNR of 0 dB, inexperienced lipreaders showed decreased word identification accuracy when the auditory signal was delayed by 300 ms relative to the visual signal. At a SNR of -10 dB, their performance declined from the synchronous condition when the auditory signal was delayed by 180 ms. For experienced lipreaders, word-identification performance deteriorated significantly by 160 ms at an SNR of -5 dB. Unfortunately, because the two groups were not tested at the same asynchrony levels or SNRs, it is difficult to form any firm conclusions about the effects of lipreading experience on integration in asynchronous AV sentences based on this study.

More recently, Grant and Seitz (1998) tested hearing-impaired adults on several measures of AV integration. They found no consistent relationship between high lipreading scores and the effects of auditory delays on AV sentence intelligibility scores. However, they also computed a measure of AV sentence benefit for synchronous sentences. This measure compared the actual improvement in AV sentence intelligibility over auditory-only sentence intelligibility with the maximum potential improvement over auditory-only sentence intelligibility. Grant and Seitz reported that AV sentence benefit scores for synchronous sentences were negatively correlated with sentence intelligibility scores at the maximum level of auditory-visual asynchrony tested. In other words, participants who were better at integrating auditory and visual information in the synchronous condition tended to have lower intelligibility scores for asynchronous auditory and visual information and thus showed greater sensitivity to the temporal coherence of auditory and visual signals. They also found that the integration score obtained from the asynchronous AV sentences was a good predictor of the AV sentence benefit score.

Although the Grant and Seitz (1998) study demonstrated a relationship between measures of AV integration and sentence intelligibility in asynchronous conditions, their study and the other studies reviewed above leave many questions unexplored. Grant and Seitz only used hearing-impaired individuals as participants. These individuals may have used fundamentally different AV integration strategies than normal-hearing individuals because of their hearing loss (see Bergeson & Pisoni, 2004). Grant and Seitz did not report correlations for raw auditory-only, visual-only, or AV sentence intelligibility scores or for several other measures of AV benefit. In addition, their study did not directly address the potential relationship between auditory, visual, or AV speech perception abilities and the simple detection of AV asynchrony in both speech and nonspeech signals. Previous research in our lab has found that several measures of AV asynchrony detection do not differ for speech or nonspeech signals (Conrey & Pisoni, this volume). Finally, only auditory-leading asynchronies were tested by both Grant and Seitz (1998) and Pandey et al. (1986), although the detection of auditory-leading asynchronies is known to be more accurate than the detection of visual-leading asynchronies (Conrey & Pisoni, 2003; Grant, van Wassenhove, & Poeppel, 2003; McGrath & Summerfield, 1985).

To begin to investigate some of these questions, we tested normal-hearing adults on their ability to detect AV asynchronies in both speech and nonspeech signals. We also measured the performance of these participants on a routine clinical sentence intelligibility task, the CUNY sentences, under auditory-only, visual-only, and AV presentation conditions. Based on previous research (Grant & Seitz, 1998; McGrath & Summerfield, 1985; Pandey, Kunov, & Abel, 1986), we hypothesized that participants with better lipreading skills and/or better AV benefit scores would be better (i.e., more accurate) at detecting AV asynchronies in both speech and nonspeech multimodal signals.

Methods

Participants

Thirty-nine Indiana University undergraduates participated in the study, which took about an hour to complete. Data from nine subjects were discarded for the following reasons: three participants did not follow directions on one or the other of the asynchrony detection task conditions (they reversed their response hand), and six participants responded "synchronous" more than 50% of the time at all asynchrony levels and their data could not be fit with the same curve-fitting procedures used for the other participants. The remaining 30 participants included 25 females and 5 males who ranged in age from 18 to 22 years (average = 19.67, *SD* = 1.03). Thirteen participants received course credit in an introductory psychology course, and 17 participants were paid $10 for their services. No significant differences between paid and unpaid participants were found on subsequent analyses, so results from the two groups were pooled in this report.

Procedure

Each participant completed the full-face (FF) and nonspeech (NS) conditions from the asynchrony detection task described in Experiment 1 of Conrey and Pisoni (2003, this volume). The participants also completed a modified version of the City University of New York (CUNY) Sentences Test (Boothroyd, Hannin, & Hnath, 1985), which is used clinically to assess auditory-only (A-only), visual-only (V-only), and auditory-visual (AV) speech perception skills in hearing impaired populations. Participants always completed the CUNY sentences first, followed by the FF and NS asynchrony detection tasks. Because the CUNY sentences are presented in the order A-only, then V-only, then AV in the clinic, they were always presented in that order in this experiment. However, the order of the FF and NS conditions of the asynchrony detection task was counterbalanced across participants. The visual stimuli were presented on an Apple Macintosh G4 computer. Auditory stimuli were presented over Beyer Dynamic DT headphones at 70 dB SPL.

Asynchrony detection. Two types of stimuli were used, speech and nonspeech. For the full-face speech (FF) condition, 10 familiar English words spoken by a female speaker of American English were chosen from the Hoosier Audiovisual Multitalker Database (Lachs & Hernandez, 1998; Sheffert, Lachs, & Hernandez, 1996). This database contains digitized AV movies consisting of single talkers speaking isolated monosyllabic words. For the nonspeech (NS) condition, the stimulus was a red circle paired with a 2000-Hz tone. On each trial, the participants were presented with an AV stimulus and were asked to indicate whether it was synchronous ("in sync") or asynchronous ("not in sync"). There were 25 levels of asynchrony, ranging at every 33 ms from the auditory leading the visual signal by 300 ms (A300V ms) to the visual leading the auditory signal by 500 ms (V500A). The stimuli were blocked by condition (FF or NS). In each condition, a participant received 10 trials at each asynchrony level, for a total of 250 trials. PsyScope version 1.5.2 (Cohen, MacWhinney, Flatt, & Provost, 1993) was used for stimulus presentation and response collection. For a more detailed description of the stimuli and the procedures used in this task, please refer to Conrey and Pisoni (this volume).

CUNY sentences. As administered clinically, the CUNY sentences consists of three sets of 12 meaningful English sentences, with one set each being presented in the A-only, V-only, then AV conditions. The observer watches and/or listens to a sentence and then is asked to repeat the sentence out loud. The clinician scores the test online based on words correctly perceived.

The use of the CUNY sentences in the present study differed in several ways from the clinical applications. First, we used normal-hearing young adults as participants. In order to avoid "ceiling effects" in the A-only and AV conditions, the auditory signal was transformed to make the test more difficult. Specifically, following the methods for locally time-reversed speech described by Saberi and Perrott (1999), the auditory signal was divided into 80 ms long segments. Each segment was time-reversed, and then the segments were recombined in their original order. Phenomenologically, this transformation produces a duplex percept in which clicking sounds and speech are perceived simultaneously. Eighty ms was chosen for the reversal interval because earlier pilot testing in our lab found that this level was more difficult than the 50 ms interval, at which participants performed at near-ceiling levels, and less difficult than the 100 ms interval, which was nearly impossible for some participants. The 80-ms intervals seemed likely to produce a range of variability in results without unduly frustrating participants.

Another difference in the present study was that the participants responded by typing their responses into the computer rather than reporting them orally to the experimenter. A pilot test of seven participants in our lab showed no significant differences between these two response modes within subjects ($F(1, 6) < 1$), and most pilot participants indicated on a questionnaire that they were more comfortable responding in the written format.

As in the clinical test, before beginning each condition—A-only, V-only, and AV—the participants were given three example trials in which they viewed and/or heard sentences but were not required to respond. During the test itself, each sentence was presented only once and the participant was asked to respond by typing what he or she thought the speaker said.

The sentences were all spoken by the same female talker from the Bernstein-Eberhardt database, and they were digitized by Theresa Hnath-Chisolm and her graduate students at the University of South Florida. In the present experiment, the sentences were presented using SuperCard running a program created in SuperEdit for the MacIntosh.

Results

Asynchrony Detection

The methods of analysis for the asynchrony detection task were as described in Conrey and Pisoni (this volume). Specifically, the proportion of times each participant responded "in sync" was calculated for every asynchrony level in the FF and NS conditions. The proportion "in sync" responses for each participant were then fitted with a Gaussian curve for the FF and NS conditions, using the program Igor Pro 4.05A Carbon (copyright 1988-2002, WaveMetrics, Inc.). The mean of the curve was taken to be the mean point of synchrony (MPS), or the point in the range of asynchrony levels that was most likely to result in a judgment of "in sync." The width of the curve, as measured by the full width at half maximum (FWHM), represented the range of asynchronies over which stimuli were judged to be synchronous more than half the time. The auditory-leading and visual-leading endpoints were the limits of the synchrony window and were calculated by subtracting (auditory-leading) or adding (visual-leading) half of the FWHM to the MPS. The averages of these measures are presented in Table 1. All statistical tests were performed on individual subject data and then averaged for descriptive purposes.

Asynchrony Conditions		Primary Measures		Derived Measures	
		MPS	FWHM	A-Lead	V-Lead
	FF	47 (15)	357 (61)	-131 (31)	225 (36)
	NS	47 (43)	400 (66)	-153 (46)	247 (61)

Table 1. Mean (standard deviation) curve fits for the AV asynchrony detection task. All numbers are in milliseconds. FF = full-face condition; NS = nonspeech condition; MPS = mean point of synchrony; FWHM = full width at half maximum; A-Lead = auditory-leading threshold; V-Lead = visual-leading threshold. "Primary measures" came directly from the curve fitting; "derived measures" were calculated based on the primary measures.

Mean Point of Synchrony (MPS). The MPSs for the FF and NS conditions were not significantly different ($F(1, 29) < 1$). This result is consistent with the earlier findings reported by Conrey and Pisoni (2003, this volume).

Full Width at Half Maximum (FWHM). The FWHM for the NS condition was significantly larger than the FWHM for the FF condition ($F(1, 29) = 7.690, p < .05$). Conrey and Pisoni (this volume) found no significant differences between the FF and NS conditions in terms of FWHM. The average difference in this study was about 28 ms, which corresponds to approximately one asynchrony level.

Auditory-leading and visual-leading thresholds. The FF and NS conditions did not differ significantly in terms of the auditory- or visual-leading thresholds of the synchrony window ($F(1, 29) = 1.547$, $F(1, 29) = 3.606$; p's > .05). These results indicate that the difference between the conditions in FWHM was relatively evenly divided between the auditory- and visual-leading sides of the synchrony window.

CUNY Sentences

Participants' responses were printed out from the output file and scored by comparison with a master list of sentences. Responses were scored using a "whole-word" method, as is typically done when the CUNY sentences are used clinically. Each word was given a score of 1 or 0 points based on whether it was completely correct or had any errors, respectively. The reversal of two letters in a word, as in teh for the, was counted as correct as long as the reversal did not form a new English word. Similarly, one-letter typographical errors that did not result in the formation of a new word were counted as correct responses.

As expected, the participants were worst overall under visual-only presentation condition, with a mean score of 16 words correct (*SD* = 7) out of 102. Auditory-only scores were next, with a mean of 44 (*SD* = 15). Participants were best at the AV condition, with a mean score of 81 (*SD* = 10). The distribution of scores for these conditions is shown in Figure 1.

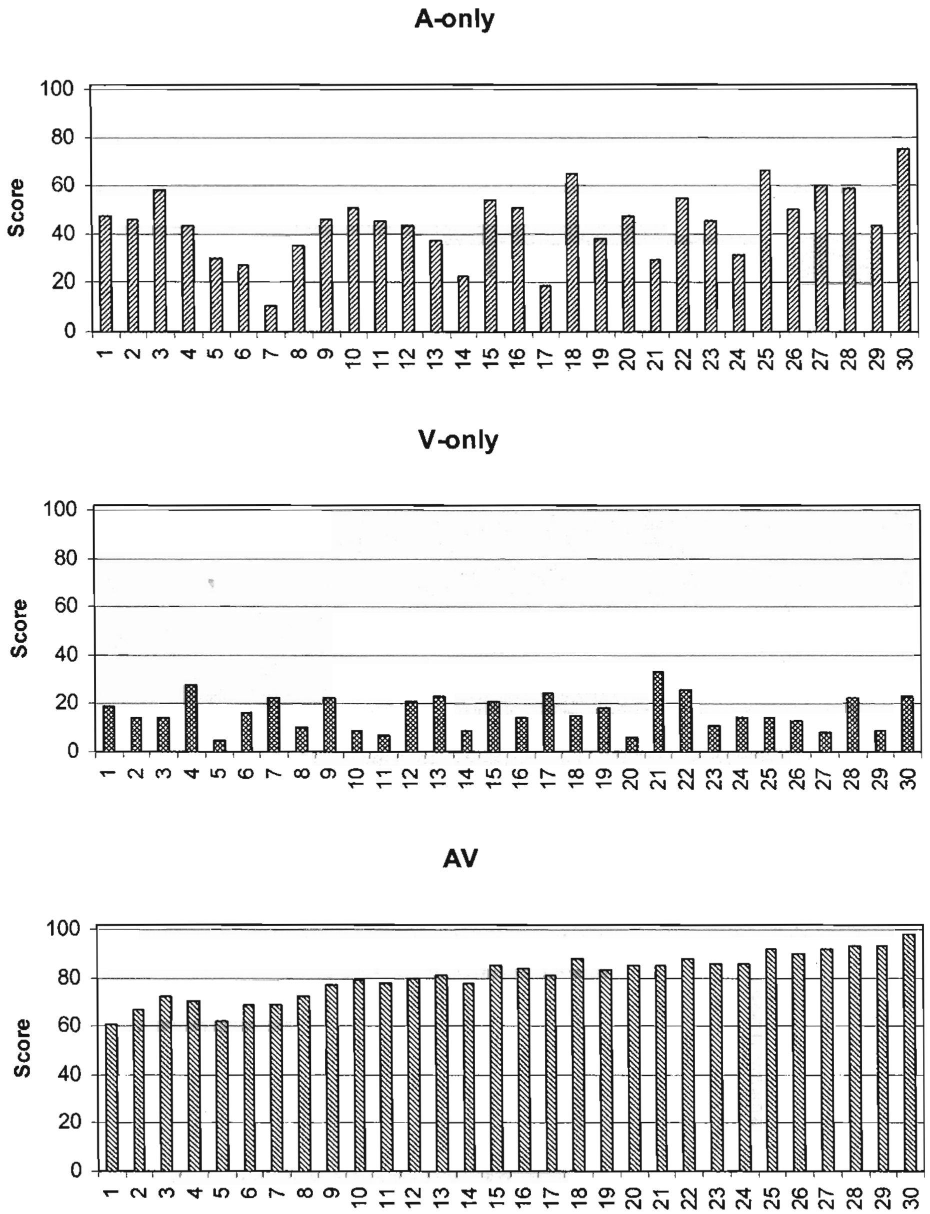

Figure 1. CUNY scores for A-only (top panel), V-only (middle panel), and AV (bottom panel) presentation conditions for each of the 30 participants, represented by participant number on the x-axis. Each score is the number of whole words correct out of a possible 102 total words.

Several other measures of performance were derived from these speech intelligibility scores. Auditory benefit was computed as the ratio of the difference between the AV score and the visual-only score (the amount the score improved after the auditory signal was added) compared with the total possible improvement in score with the added auditory signal. This measure was calculated as (AV - V) / (1 - V). Visual benefit was computed as the ratio of the difference between the AV and auditory-only scores compared with the total possible improvement in score with the added visual signal. This measure was calculated as (AV – A) / (1 – A). Finally, a measure of AV gain—the actual benefit resulting from the presence of both auditory and visual signals compared with the benefit expected from the simple sum of the scores for the individual modalities—was calculated as AV / (A + V). This measure and other similar ones have been employed in animal studies of multisensory integration (Stein & Meredith, 1993) and also in research on the BOLD response in fMRI studies (Calvert, Hansen, Iversen, & Brammer, 2001). If the ratio is greater or less than 1, then a supra-additive or subadditive effect has occurred; these types of effect have served as signatures of multisensory interactions in the literature.

Correlational Analyses

Visual inspection of the individual data revealed that some of the measures used were not normally distributed, and so correlational analyses were performed using Spearman's rho (r_s) rather than Pearson's r.

Asynchrony detection. Among the measures obtained directly from the curve fits, the FWHMs were positively correlated for the FF and NS conditions, $r_s = +.65$, $p < .01$. Also, the MPSs were positively correlated for the FF and NS conditions, $r_s = +.46$, $p < .01$.

Among the measures derived from the MPS and FWHM, the visual-leading thresholds were positively correlated for the FF and NS conditions, $r_s = +.50$, $p < .01$. Also, the FF auditory-leading and visual-leading thresholds were negatively correlated, $r_s = -.47$, $p < .05$. Because auditory-leading thresholds were coded as negative numbers and visual-leading thresholds were coded as positive, this finding indicates that larger ("lower") auditory-leading thresholds were related to larger ("higher") visual-leading thresholds. The relationship between auditory- and visual-leading thresholds for the NS condition was not significant, however ($r_s = +.17$, $p > .05$). Several of the correlations between "primary" and "derived" asynchrony detection performance measures were significant, although this finding may not have much practical significance because the derived measures were calculated using the primary measures. A summary of the intercorrelations is given in Table 2.

			Primary				Derived			
			FF		NS		FF		NS	
			MPS	FWHM	MPS	FWHM	A-Lead	V-Lead	A-Lead	V-Lead
Primary	FF	MPS	—							
		FWHM	.18	—						
	NS	MPS	.46**	.30	—					
		FWHM	-.04	.65***	.24	—				
Derived	FF	A-Lead	.28	-.85***	-.01	-.68***	—			
		V-Lead	.65***	.82***	.47*	.41*	-.47**	—		
	NS	A-Lead	0.37*	-.12	.67***	-.47*	.37*	.15	—	
		V-Lead	.35	.47**	.81***	.68***	-.31	.51**	.17	—

Table 2. Correlations among measures of AV asynchrony detection. Abbreviations as in Table 1.
* = $p < .05$; ** = $p < .01$; *** = $p < .001$.

CUNY sentences. A-only and AV scores were positively correlated, $r_s = +.53$, $p < .01$. The correlations between V-only and A-only or AV scores were small and not significant. Auditory benefit and visual benefit were positively correlated, however ($r_s = +.85$, $p < .001$). In addition, A-only scores and V-only scores were negatively correlated with AV gain ($r_s = -.71$, $r_s = -.42$; $p < .001$, $p < .05$, respectively). These correlations indicate that larger AV gains were associated with lower unimodal scores. Both A-only and AV scores were positively correlated with auditory benefit ($r_s = +.56$, $r_s = +.97$, respectively; p's < .001), and AV score was also positively correlated with visual benefit, $r_s = +.89$, $p < .001$. Table 3 shows the bivariate correlations among the primary and derived measures from the CUNY sentences task.

		Primary			Derived		
		A-Only	V-Only	AV	A-Ben	V-Ben	AV Gain
Primary	A-Only	---					
	V-Only	-.06	---				
	AV	.53**	.03	---			
Derived	A-Ben	.56***	.10	.97***	---		
	V-Ben	.16	-.15	.89***	.85***	---	
	AV Gain	-.71***	-.42*	-.04	.01	.27	---

Table 3. Correlations among measures of AV sentence intelligibility (CUNY sentences). A-Ben = auditory benefit, V-Ben = visual benefit. "Derived" means derived from primary scores. * = $p < .05$; ** = $p < .01$; *** = $p < .001$.

Asynchrony detection and CUNY sentences. Several of the intercorrelations among the asynchrony detection and CUNY sentences measures were also significant. AV scores were negatively correlated with the FWHM for both the FF and NS conditions ($r_s = -.47$, $r_s = -.45$; $p < .01$, $p < .05$, respectively). A-only scores were also negatively correlated with the FWHM for the NS condition, $r_s = -.41$, $p < .05$. These results indicate that higher AV scores for the FF and NS conditions and higher A-only scores for the NS condition were associated with smaller synchrony windows, or less tolerance for asynchrony. In addition, auditory benefit was correlated with the FWHM for both the FF and NS conditions ($r_s = -.48$, $r_s = -.52$, respectively; p's < .01), and visual benefit was negatively correlated with the FWHM for the FF condition ($r_s = -.39$, $p < .05$). Finally, AV, A-only, auditory benefit, and visual benefit scores were all positively correlated with the auditory-leading threshold for the FF condition ($r_s = +.49$, $r_s = +.39$, $r_s = +.55$, $r_s = +.43$, respectively; p's < .05). This pattern indicates that higher AV and A-only scores and higher multimodal benefit scores for sentence intelligibility were associated with lower auditory-leading thresholds that were closer to physical synchrony. Table 4 summarizes these results.

				CUNY Sentences					
				Primary			Derived		
				A-Only	V-Only	AV	A-Ben	V-Ben	AV Gain
Asynchrony Detection	Primary	FF	MPS	.10	-.29	.08	.15	.08	.19
			FWHM	-.31	-.04	-.47**	-.48**	-.39*	.07
		NS	MPS	-.12	-.17	-.09	-.02	-.02	.23
			FWHM	-.41*	.20	-.45**	-.52**	-.35	.06
	Derived	FF	A-Lead	.39*	-.14	.49**	.55**	0.43*	.00
			V-Lead	-.10	-.15	-.31	-.27	-.28	.05
		NS	A-Lead	.14	-.17	.21	.30	.21	.08
			V-lead	-.31	.14	-.14	-.18	-.12	.21

Table 4. Correlations among measures of AV asynchrony detection and AV sentence intelligibility. Abbreviations as in previous tables. * = p < .05; ** = p < .01; *** = p < .001.

Discussion

The results of this study demonstrate several relations between measures of AV sentence intelligibility and the ability to detect AV asynchrony in both speech and nonspeech signals. Specifically, the participants who obtained higher AV sentence intelligibility scores tended to have smaller windows over which they identified AV signals as synchronous and thus were more accurate at detecting small differences in AV asynchrony in speech and nonspeech signals. In addition, higher auditory and visual sentence benefit scores were also correlated with smaller AV synchrony windows for both speech and nonspeech signals. Finally, higher auditory-only, AV, and auditory and visual benefit scores were correlated with auditory-leading thresholds in the FF condition that were closer to physical synchrony. This pattern was not observed in the NS condition. This result implies that subjects who performed better on AV sentence intelligibility measures were more accurate at identifying the asynchrony in auditory-leading speech, but not in auditory-leading nonspeech signals. This finding is interesting given that the auditory signal typically lags behind the visual signal in natural speech, because the articulators must be positioned before the sound can be made. Perhaps the better AV integrators were able to take advantage of this natural relationship in speech, whereas the lack of expectation about which signal should lead in the artificial nonspeech signal condition prevented the use of natural cues in that condition.

Several researchers have suggested that AV asynchrony detection may have its neural basis in multisensory processing times (Conrey & Pisoni, this volume; Lewald, Ehrenstein, & Guski, 2001). However, in earlier research it was unclear whether AV asynchrony detection was meaningfully related to AV integration measures relevant for speech perception. The results of the present study have revealed strong relations among measures of AV integration and the detection of AV asynchrony in speech and nonspeech signals that provide several new insights into the common cognitive and neural mechanisms relevant for both multisensory temporal sensitivity and speech perception, especially perception of multimodal speech signals. Fundamental differences in neural timing of multimodal sensory events may be linked to the variation observed in lipreading performance and multimodal benefit among normal-hearing individuals.

References

Bergeson, T. R., & Pisoni, D. B. (2004). Audiovisual speech perception in deaf adults and children following cochlear implantation. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *Handbook of multisensory processes* (pp. 749-772). Cambridge, MA: MIT Press.

Boothroyd, A., Hannin, L., & Hnath, T. (1985). A sentence test of speech perception: Reliability set equivalence and short-term learning (internal report RCI 10). New York: City University of New York.

Calvert, G., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage, 14,* 427-438.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*, 257-271.

Conrey, B. L., & Pisoni, D. B. (2003). Audiovisual asynchrony detection for speech and nonspeech signals. Proceedings of the Audio Visual Speech Processing (AVSP) Workshop, 25-30.

Conrey, B. L., & Pisoni, D. B. (this volume). Detection of auditory-visual asynchrony in speech and nonspeech signals. In *Research on Spoken Language Processing Progress Report No. 26* (pp. 71-94). Bloomington, IN: Speech Research Laboratory, Indiana University.

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

Grant, K.W., van Wassenhove, V. & Poeppel, D. (2003). Discrimination of auditory-visual synchrony. Proceedings of the Audio Visual Speech Processing (AVSP) Workshop, 31-35.

Lachs, L., & Hernandez, L. R. (1998). Update: The Hoosier audiovisual multitalker database. In *Research on Spoken Language Processing Progress Report No. 22* (pp. 377-388). Bloomington, IN: Speech Research Laboratory, Indiana University.

Lewald, J., Ehrenstein, W.H., & Guski, R. (2001). Spatio-temporal constraints for auditory-visual integration. *Behavioral Brain Research, 121*, 69-79.

McGrath, M. & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults, *Journal of the Acoustical Society of America*, *77*, 678-684.

Pandey, C.P., Kunov, H. & Abel, M.S. (1986). Disruptive effects of auditory signal delay on speech perception with lip-reading, *The Journal of Auditory Research*, *26*, 27-41.

Saberi, K., & Perrott, D. R. (1999). Cognitive restoration of reversed speech. *Nature, 398,* 760.

Sheffert, S. M., Lachs, L., & Hernandez, L. R. (1996). The Hoosier audiovisual multitalker database. In *Research on Spoken Language Processing No. 21* (pp. 578-583). Bloomington, IN: Speech Research Laboratory, Indiana University.

Stein, B. & Meredith, M.A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.