Effects of Presentation Mode and Repeated Familiarization on Intelligibility of Dysarthric Speech

Katherine C. Hustad
Meghan A. Cahill
The Pennsylvania State University, University Park

Clinical measures of speech intelligibility are widely used as one means of characterizing the speech of individuals with dysarthria. Many variables associated with both the speaker and the listener contribute to what is actually measured as intelligibility. The present study explored the effects of presentation modality (audiovisual vs. audio-only information) and the effects of speaker-specific familiarization across 4 trials on the intelligibility of speakers with mild and severe dysarthria associated with cerebral palsy. Results revealed that audiovisual information did not enhance intelligibility relative to audio-only information for 4 of the 5 speakers studied. The one speaker whose intelligibility increased when audiovisual information was presented had the most severe dysarthria and concomitant motor impairments. Results for speaker-specific repeated familiarization were relatively homogeneous across speakers, demonstrating significant intelligibility score improvements across 4 trials and, in particular, a significant improvement in intelligibility between the 1st and 4th trials.

Key Words: cerebral palsy, speech intelligibility, dysarthria, listener learning

Speech intelligibility is an important construct in the assessment and management of individuals who have dysarthria. Intelligibility scores are often used for clinical purposes such as quantifying severity of dysarthria and measuring progress before, during, and after intervention (Hustad, Beukelman, & Yorkston, 1998; Yorkston, Beukelman, Strand, & Bell, 1999). Indeed, important decisions such as termination or continuation of treatment and consideration of augmentative and alternative communication systems may be based, in part, on intelligibility scores. As such, a comprehensive understanding of the factors that influence intelligibility and its measurement is critical.

Speech intelligibility has been broadly defined as the accuracy with which a listener is able to decode the acoustic signal of a speaker (Kent, Weismer, Kent, & Rosenbek, 1989; Yorkston & Beukelman, 1980; Yorkston, Strand, & Kennedy, 1996). It is an interactive process (Connolly, 1986; Hustad et al., 1998; Kent, 1993; Yorkston, Strand, et al., 1996) that fluctuates for any given speaker depending on a host of variables pertaining not only to the acoustic signal produced by the speaker, but also to variables associated with the listener and the communicative context (Kent et al., 1989).

Relatively few of the variables thought to influence intelligibility have been systematically studied. However, research has begun to demonstrate the independent and powerful impact on intelligibility of what Lindblom (1990) referred to as "signal-independent information", or variables extrinsic to the acoustic signal. For example, semantic predictability dramatically affects listeners' ability to understand dysarthric speech, with sentences that are more predictable resulting in higher intelligibility scores than sentences that are less predictable (Garcia & Cannito, 1996; Garcia & Dagenais, 1998). In addition, supplemental linguistic contextual information such as alphabet cues, topic cues, and combined alphabet and topic cues have been shown to increase intelligibility of severely dysarthric speech substantially (Hustad & Beukelman, 2001). Similarly, provision of iconic gestures concurrently with speech can significantly enhance intelligibility (Garcia & Cannito, 1996). Other variables that may affect intelligibility include presentation mode of the signal (audiovisual vs. audio only) and familiarization with the speech signal. However, research examining each of these two variables as they relate to individuals with dysarthria has been equivocal, as reviewed below, suggesting the need for further research.

1058-0360/03/1202-0198

Effects of Visual Information on Intelligibility

During face-to-face communication interactions, speaker and listener share the benefit of visual information such as phoneme place cues, facial expression, body language, and gestures. These visual cues provide the listener with additional information that supplements and may add redundancy to the acoustic signal (Erber, 1975). In general, research suggests that several regions of the face contribute to visual speech perception (Munhall & Vatikiotis-Bateson, 1998; Preminger, Lin, Payen, & Levitt, 1998). In particular, visible phonetic information is provided by the lips, oral aperture, and mandible (Munhall & Vatikiotis-Bateson, 1998) and may serve to provide listeners with information that aids in decoding speech.

Research examining the effects of visual information on intelligibility of speech produced by non-motor-impaired speakers has consistently shown that intelligibility of frequency-filtered speech (Sanders & Goodrich, 1971), hearing-impaired speech (Erber, 1975; Monsen, 1983), and esophageal speech (Hubbard & Kushner, 1980) is higher when listeners receive simultaneous audiovisual information relative to the auditory signal in isolation. However, few studies have examined the effects of audiovisual versus audio-only presentation modes on intelligibility of motor-impaired speakers who have dysarthria. Results of existing studies are contradictory, with some research demonstrating marked increases in intelligibility associated with audiovisual input and others showing no difference between audiovisual and audio-only input.

Existing research seems to suggest that severity of dysarthria may play a role in determining whether there is an intelligibility benefit for audiovisual versus audio-only information. For speakers with severe dysarthria, descriptive data from two studies have suggested that intelligibility differences between auditory and auditory–visual presentation modes are not clinically meaningful (up to 3% difference; Garcia & Cannito, 1996; Hunter, Pring, & Martin, 1991). However, another study found contradictory results for 2 speakers with severe dysarthria, demonstrating a much larger increase in intelligibility (between 8% and 18%) when listeners received audiovisual input (Garcia & Dagenais, 1998). For speakers with moderate dysarthria, research is somewhat more consistent, suggesting that intelligibility is markedly improved when listeners receive audiovisual versus audio-only input (between 11% and 17%; Garcia & Dagenais, 1998; Hunter et al., 1991). To date, there has been no research examining the effects of presentation modality on intelligibility of mild dysarthria.

Understanding the effects of visual information on intelligibility of speakers with dysarthria is necessary in order to ensure the ecological validity of clinical assessment. Standard clinical measures such as the Sentence Intelligibility Test (SIT; Yorkston, Beukelman, & Tice, 1996) or the Assessment of Intelligibility of Dysarthric Speech (Yorkston & Beukelman, 1981) involve the use of audio-recorded speech samples, which may not accurately represent intelligibility scores in face-to-face communication situations if audiovisual presentation consistently enhances intelligibility. Additional research is necessary to further clarify the effects of presentation modality on intelligibility, particularly for individuals with mild and severe dysarthria.

Effects of Familiarization on Intelligibility

The effects of listener familiarization on intelligibility have been studied for several populations including individuals who have hearing impairment, dysarthria, and synthesized or digitally manipulated speech signals. Studies examining the effects of familiarization have generally revealed that listener intelligibility scores improve following repeated exposure to specific speakers or voices (Hoover, Reichle, Van Tasell, & Cole, 1987; McNaughton, Fallon, Tod, Weiner, & Neisworth, 1994; Rosen, Faulkner, & Wilkinson, 1999; Schwab, Nusbaum, & Pisoni, 1985; Spitzer, Liss, Caviness, & Adler, 2000; Tjaden & Liss, 1995; Venkatagiri, 1994). However, the few studies examining familiarization with dysarthric speech have, again, revealed some contradictory results.

Tjaden and Liss (1995) provided speaker-specific familiarization training in which listeners followed a written transcript of target productions concurrently with presentation of the acoustic signal for one individual with moderate dysarthria. Transcription results for different stimuli produced by the same speaker revealed that sentence intelligibility improved significantly for listeners who received familiarization training relative to a control group of listeners who did not receive familiarization training. Garcia and Cannito (1996) also provided a speaker-specific familiarization experience to listeners for one speaker with severe dysarthria. However, they did not provide listeners with a written transcript to follow while viewing the familiarization sample. Their results showed that there was no significant difference in intelligibility scores between listeners who were familiarized and those who were not. The discrepancy between these two studies may be related to different familiarization procedures or individual differences among the speakers studied.

Studies examining larger groups of speakers have also shown inconsistent findings. Spitzer and colleagues (2000) found that when listeners were provided with a written transcript of target productions concurrently with presentation of the acoustic signal for two groups of speakers (those with moderately severe ataxic dysarthria and those with moderately severe hypokinetic dysarthria), intelligibility increased significantly between pre- and postfamiliarization measures for both groups. The overall familiarization effect observed in this study is consistent with research examining familiarization with synthesized voices (Hoover et al., 1987; McNaughton et al., 1994; Schwab et al., 1985; Venkatagiri, 1994). However, in a study examining the effects of three familiarization procedures on intelligibility of dysarthric speech of varying severity, Yorkston and Beukelman (1983) found results that contradicted those of Spitzer and colleagues. In this study, intelligibility scores were examined for each of three listener groups: those who heard samples of each speaker repeated twice, those who heard samples of each speaker three times while following along with a written transcript containing the target sentences, and those who did not receive any familiarization experience. Results showed that there was no difference between pre- and postfamiliarization intelligibility measures for any of the three groups. This discrepancy in findings between Spitzer and colleagues and Yorkston and Beukelman (1983) is difficult to explain, particularly considering that similar methods were used. However, one possible explanation is insufficient statistical power for the Yorkston and Beukelman (1983) study, which involved a total of 9 listeners (3 per condition). The study by Spitzer and colleagues involved a total of 60 listeners (30 per group) and, consequently, their findings are likely to be more statistically robust.

Existing studies examining the effects of familiarization on intelligibility of dysarthric speech have had only two repeated measures, pre- and posttests. In addition, most of these studies have used explicit structured familiarization experiences. The effects of repeated familiarization experiences (beyond pre- and posttests) in which instruction and/or feedback is not provided have not been studied, but they have important implications for speakers with dysarthria. If repeated listening experience results in meaningful intelligibility gains, one intervention strategy to enhance the effectiveness of dysarthric speech may involve simple repetitive familiarization experiences for listeners.

The present study examined the effects of presentation modality (audiovisual vs. audio-only input) and the effects of four familiarization trials on the intelligibility of 5 speakers with mild and severe dysarthria. The following specific research questions were addressed:

1. What is the effect of audiovisual versus audio-only input on overall intelligibility scores and individual intelligibility scores for 5 speakers with dysarthria?
2. What is the effect of four repeated familiarization trials on overall intelligibility scores and individual intelligibility scores for 5 speakers with dysarthria?

Method

Speakers With Dysarthria

Five individuals with mild or severe dysarthria secondary to cerebral palsy (CP) served as speakers. All indicated that they used speech as their primary mode of communication. Participation required that speakers meet the following criteria: (a) speech intelligibility between 70% and 95% (mild dysarthria) or between 10% and 35% (severe dysarthria) as measured by the SIT (Yorkston et al., 1996), (b) native speakers of American English, (c) between 18 and 50 years of age, (d) able to produce connected speech consisting of at least eight consecutive words, and (e) able to repeat sentences of up to eight words in length following a verbal model. Demographic data on speakers, including type and severity of dysarthria (determined by a certified speech-language pathologist) and type of CP (based on medical diagnosis) are presented in Table 1.

Speech Stimuli

This study used sentence stimuli from the Hearing in Noise Test (HINT; Nilsson, Soli, & Sullivan, 1994). The HINT consists of 25 lists each containing 10 sentences that are phonemically balanced and equated for naturalness, difficulty, length, and reliability (Nilsson et al., 1994). Sentences within each list are declarative in nature, follow a predictable subject–verb–object syntactic structure, and range in length between four and seven words.

For the present study, HINT Lists 1–4 were used. Because predictiveness of sentences was not formally controlled in the development of the HINT, potential predictiveness differences among lists were controlled through quasi-counterbalancing (as described below). This assured that predictiveness was approximately uniformly distributed across listeners and other independent variables.

TABLE 1. Demographic data for individual speakers with dysarthria.

Demographic category	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5
Age (years)	34	37	46	50	24
Gender	M	M	F	M	F
Medical diagnosis	Athetoid diplegia	Spastic diplegia	Spastic diplegia	Spastic diplegia	Athetoid quadriplegia
Type of dysarthria	Hyperkinetic	Spastic	Spastic	Spastic	Mixed spastic–hyperkinetic
Education	College graduate	Currently working on GED	Attended college on a part-time basis for 5 years	College graduate	Currently attending community college
Intelligibility on SIT (%)	80	75	30	24	16

Recording Speech Samples

Speakers were required to produce the full corpus of 40 HINT sentences. Digital audio and video recordings of each speaker were made in a double-walled soundproof room. In addition to the audio signal recorded by the digital video (DV) camera, separate digital audiotape (DAT) recordings were made to allow manual adjustment and monitoring of sound levels. DAT samples were recorded using an HHb PDR1000 Portadat recorder with a sampling rate of 44.1 kHz, high pass filtering at 100 Hz, and a Crown CM-312 head-mounted microphone positioned 5 cm from the speaker's mouth. The microphone was positioned so that it was aligned with the jawline of the speaker and did not obscure any aspect of the speaker's face or mouth. DV samples were recorded on a Sony TRV 900 miniDV camcorder that was positioned approximately 3 feet away from and directly in front of each speaker, focused inferiorly at mid-torso and superiorly just above the speaker's head. All speakers were seated directly in front of the same pale colored neutral background with the same external lighting positioned to eliminate shadows during recording. Speakers produced each sentence following the experimenter's model. Orthographic representations of stimulus sentences were provided on a computer screen placed in front of the speaker, yet out of video camera view. Speakers were instructed to speak "naturally," as they would in habitual communication situations. Rate and prosody for each speaker were not controlled.

Stimulus Tape Preparation

Audio recordings were transferred from DAT to a personal computer via a digital-to-digital sound card (S/PDIF interface; 16-bit quantization; 44.1-kHz sampling rate), using Sound Forge 4.5 (Sonic Foundry, 1998) computer software. Individual sentences were edited and peak amplitude normalized to 69 dB. Final signal-to-noise ratios were above 45 dB for all recordings.

Videorecordings were transferred to a personal computer via a Firewire (IEEE 1394) card, maintaining the sampling rate and frame size of the original DV recording (video = 29.97 frames per second, 720 × 480 frame size). Audio files from DAT were then matched with the digital video files via Adobe Premiere 5.1 (digital video software). Visual inspection of the two waveforms (original from DV camera and higher quality from DAT) and auditory–perceptual judgments obtained by listening to the two audio samples simultaneously were used to ensure perfect alignment. Following alignment of the two samples, the original audio sample was deleted, leaving only the high quality DAT sample associated with the video of each stimulus sentence.

Files for each stimulus sentence within each HINT list were then sequenced as follows: (1) written instructions for the trial, (2) sentence number, (3) target sentence, (4) sentence number, (5) target sentence. Items 2–5 were repeated until each of the 10 HINT sentences was presented. Following this first presentation of all 10 stimulus sentences, the following items were presented: (1) written instructions directing listeners to write down what they heard during the interval following each sentence, (2) sentence number, (3) target sentence, (4) written instructions to transcribe the preceding sentence, (5) sentence number, (6) target sentence, (7) written instructions to transcribe the preceding sentence. Items 2–7 were repeated until all 10 sentences were completed. The final sequence was then exported from the computer to digital videotape, maintaining first generation quality audio and video signals following the National Television System Committee broadcast-quality standards.

For tapes containing the audio-only stimuli, files made for each speaker and each HINT list, as described above, were edited so that video images of the speaker were digitally removed and replaced by a plain blue screen. Consequently, trial duration, instructions, and pauses were identical for audio-only and audiovisual conditions, with the only difference being the replacement of the video images with a plain blue screen. Files for the audio-only condition were also exported to digital videotape and presented following the same procedures as in the audiovisual condition.

Listeners

Twenty different listeners were randomly assigned to each of the 5 speakers, for a total of 100 listeners. Listeners met the following criteria: (a) pass a pure tone hearing screening at 25 dB SPL for 250 Hz, 500 Hz, 1000 Hz, 4000 Hz, and 6000 Hz bilaterally; (b) between 18 and 30 years old; (c) no more than incidental experience listening to or communicating with persons having communication disorders; (d) native speakers of American English; and (e) no identified language, learning, or cognitive disabilities per self-report. Listeners within each speaker group were drawn primarily from a pool of college students. The mean age of listeners in each group ranged from 20 to 21 years. Because gender was not a variable of interest, no effort was made to balance the number of male and female listeners.

Randomization and Counterbalancing

Listeners participated in two trials that involved audio-only presentation and two trials that involved audiovisual presentation, for a total of four experimental trials. Trials for each of the two modes of presentation always occurred adjacent to one another so that half of the listeners in each speaker group completed two audio-only trials followed by two audiovisual trials, and the other half of the listeners in each speaker group completed two audiovisual trials followed by two audio-only trials. Consequently, within each of the four experimental trials, half of the data for each of the five speaker groups were associated with the audio-only condition and half were associated with the audiovisual condition. Data addressing the effects of familiarization reflect averages across these presentation modes.

Listeners heard a different HINT list for each experimental trial. To ensure equivalency of lists, presentation order of lists within each speaker group was quasi-counterbalanced, with every listener receiving the four HINT lists in a different order.

Experimental Task

Presentation of Stimuli. The digital audiovideo signal was presented to individual listeners in a sound-treated quiet listening environment. During the experiment, the listeners were seated at a desk and positioned approximately 3 feet away from a 25-in. television monitor with one external speaker and a digital video cassette player attached to it. The television monitor was positioned at listeners' eye level (approximately 3.5 feet from the ground). The external speaker was positioned at listeners' chest level (approximately 2.5 feet from the ground), directly below the television monitor. The peak output level of stimulus material was approximately 65 dB SPL from where listeners were seated and was measured periodically to ensure that all listeners heard stimuli at the same output level.

Administration Instructions. Listeners were told that they would complete four experimental trials that would take a total of approximately 50 min. The experimenter explained that for each of the trials, they would see and/or hear the same person who has a speech impairment because of cerebral palsy. Following delivery of specific trial instructions (provided in the Appendix), listeners were given the opportunity to ask questions. After providing answers to any questions the listeners may have asked, the experimenter left the sound-treated room and entered an adjacent control room from which she controlled the digital video player and observed the listeners. After each trial, the experimenter entered the sound-treated room to change the videotape.

Scoring and Reliability

Intelligibility for each listener was scored as the number of words identified correctly divided by the number of words possible for each trial. Individual words were judged as incorrect or correct on the basis of whether they matched the target word phonemically. Misspellings and homonyms were accepted as correct.

Interscorer reliability involved having a judge, who was not involved in initial scoring of intelligibility data, rescore all transcription data for 10 of the 100 listeners (2 listeners from each speaker group). The original transcription results (in percent intelligibility) for the same listener across each of the four lists were then correlated with the rescored transcription results, yielding a Pearson product–moment correlation coefficient of .99 across all 10 listeners, indicating a very high level of reliability for scoring accuracy.

Experimental Design and Analysis

To address the research questions of interest, it was necessary to conduct two separate analyses of variance (ANOVA). Each ANOVA provided a different view of the experimental data and used a different statistical model. Thus, analyses could not be integrated into one fully factorial analysis.

The first set of analyses examined experimental data collapsed across speakers and familiarization trials and used a 2×2 split-plot design. The repeated measure was presentation mode and its two categories were audio-only and audiovisual presentation. The between-subjects measure was presentation order, with one group receiving the audio-only conditions first and one group receiving the audiovisual conditions first. This ANOVA used an interaction model, with the Type I error rate set at .01 for each omnibus test.

The second ANOVA examined experimental data separately for each speaker, using a 2×4 repeated-measures design (Kirk, 1995). The two-level repeated measure was presentation mode and its categories were audio-only and audiovisual presentation. The four-level repeated measure was familiarization trials (across presentation modes) and its categories were first, second, third, and fourth. This ANOVA used a nested model (Kirk, 1995; Marascuilo & Serlin, 1988) and the Type I error rate was set at .01 for each omnibus test. Follow-up contrasts associated with each omnibus test were allotted a familywise alpha of .05, which was partitioned using the Dunn Bonferroni procedure (Kirk, 1995; Marascuilo & Serlin, 1988), yielding a per contrast alpha of .008.

Results

Analysis 1

The mean intelligibility score across all speakers for the audiovisual mode of presentation was 63.74% (SD = 25.09%). For the audio mode of presentation, mean intelligibility was 59.69% (SD = 25.74%). ANOVA results indicated that this main effect for presentation mode was statistically significant. See Table 2 for specific ANOVA statistics.

The mean intelligibility score across all speakers for the group of listeners that received the audiovisual trials first was 62.88% (SD = 49.71%). For the group of listeners that received the audio-only trials first, mean intelligibility was 60.82% (SD = 49.71%). ANOVA results indicated that this main effect for presentation order was not statistically significant, indicating that the two groups were equivalent.

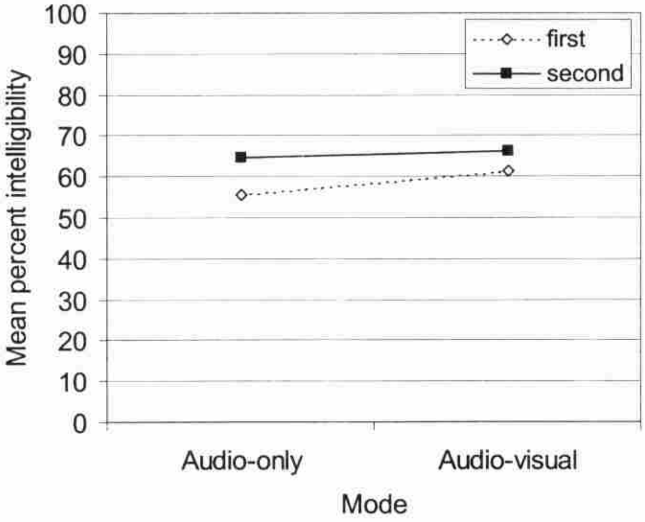
Finally, ANOVA results revealed that the interaction between presentation mode and presentation order was statistically significant, indicating that the difference

TABLE 2. Interaction model ANOVA results for Analysis 1.

Source	df	F	η^2
		Between subjects	
Presentation order	1	0.171	.002
Error	98	(211.758)	
		Within subjects	
Mode	1	12.606*	.114
Mode × Presentation Order	1	43.705*	.308
Error	98	(56.551)	

Note. Values enclosed in parentheses represent mean square errors.
*$p < .001$.

FIGURE 1. Interaction between presentation mode and presentation order.



between audio-only and audiovisual modes was greater for listeners who received the audio condition first (a 10% increase in intelligibility between modes) than for those who received the audiovisual condition first (a 3% increase in intelligibility between modes). See Figure 1 for a graphic display of this interaction. Because presentation order was systematically varied for each speaker group, this interaction effect was evenly distributed within each speaker group and, therefore, did not result in a differential influence on learning data averaged across modes.

Analysis 2

Mode Within Speakers. Nested omnibus tests examining mode within each speaker group showed that the difference between audio-only and audiovisual presentation modes was not significant for Speakers 1, 2, 3, and 4. However, for Speaker 5, this difference was significant, with a mean difference of 10.18% in favor of the audiovisual presentation mode. Individual speaker data are displayed graphically in Figure 2. Specific statistics for this ANOVA are presented in Table 3.

Familiarization Trials Within Speakers. Mean intelligibility scores across all speakers for each presentation, regardless of mode, were as follows: 55.49% (SD = 26.79%) for the first presentation, 61.18% (SD = 27.93%) for the second presentation, 64.25% (SD = 25.20%) for the third presentation, and 66.48% (SD = 24.11%) for the fourth presentation. ANOVA results revealed a significant main effect of familiarization trials, and follow-up contrasts indicated that there was a significant linear trend. In addition, across all speakers, intelligibility scores on the first presentation were worse than intelligibility scores on the second, the third, and the fourth presentations. Intelligibility scores on the fourth presentation were significantly better than intelligibility scores on the second presentation. Finally, intelligibility scores on the third presentation did not differ from those on the second or the

FIGURE 2. Mean intelligibility scores (±1 *SD*) by presentation mode (audio-only and audio-visual) and speaker.

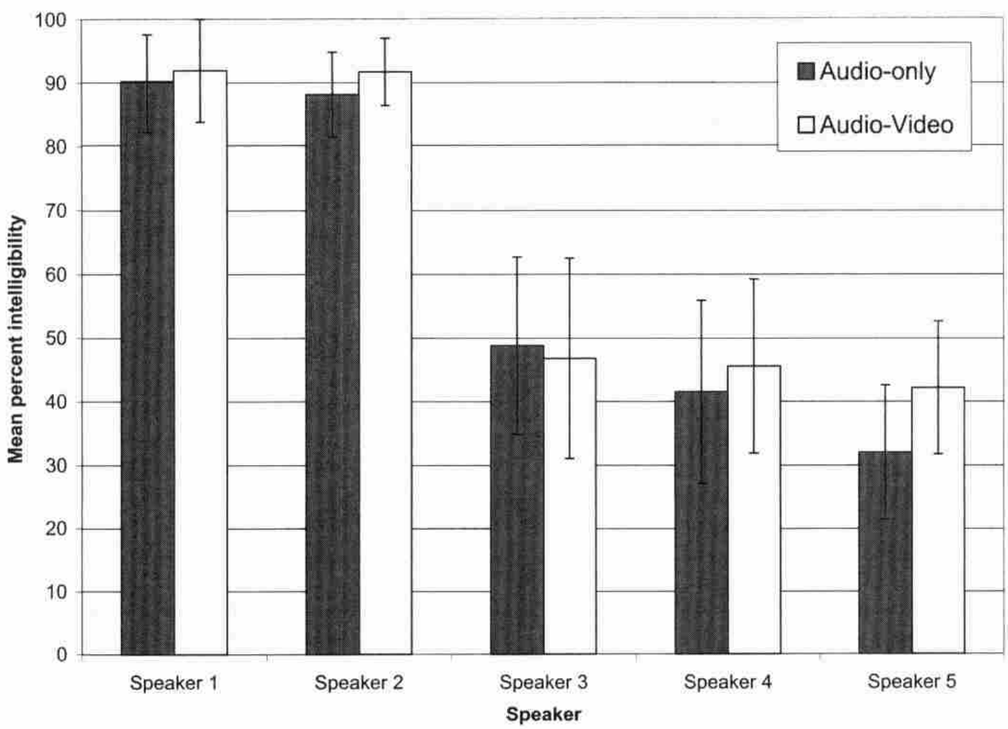


TABLE 3. Nested model ANOVA results for Analysis 2.

Source	*df*	*F*	η^2
Within subjects			
Mode (across speakers)	1	12.606*	.114
Error	98	(56.551)	
Mode nested in speaker			
Speaker 1 (mode)	1	0.914	.046
Error	19	(61.218)	
Speaker 2 (mode)	1	4.931	.206
Error	19	(52.568)	
Speaker 3 (mode)	1	0.339	.018
Error	19	(259.911)	
Speaker 4 (mode)	1	1.076	.054
Error	19	(308.718)	
Speaker 5 (mode)	1	15.470*	.450
Error	19	(131.884)	
Within subjects			
Familiarization (across speakers)	2.714	25.828*	.207
Error	268.677	(97.217)	
Familiarization nested in speaker			
Speaker 1 (trials)	1.903	8.535*	.310
Error	36.153	(69.085)	
Speaker 2 (trials)	2.199	7.233**	.276
Error	41.780	(38.884)	
Speaker 3 (trials)	2.689	4.361***	.187
Error	51.084	(165.644)	
Speaker 4 (trials)	2.418	6.648**	.259
Error	45.938	(156.942)	
Speaker 5 (trials)	2.157	5.957**	.239
Error	40.992	(128.809)	

Note. Values enclosed in parentheses represent mean square errors.
$^*p < .001$. $^{**}p < .005$. $^{***}p < .01$.

fourth presentations. Statistics for follow-up contrasts are presented in Table 4.

Data for Speaker 1 showed that mean intelligibility scores, regardless of presentation mode, were as follows: 84.78% (*SD* = 9.94%) for the first presentation, 93.72% (*SD* = 4.48%) for the second presentation, 92.93% (*SD* = 5.29%) for the third presentation, and 93.55% (*SD* = 6.12%) for the fourth presentation. This omnibus ANOVA was significant. Consequently, six follow-up contrasts were examined. Results showed that intelligibility scores on the first presentation were worse than intelligibility scores on the second, the third, and the fourth presentations. No other pairwise contrasts were significant.

Data for Speaker 2 revealed the same pattern of results as those obtained for Speaker 1. Mean intelligibility scores, regardless of presentation mode were as follows: 85.20% (*SD* = 7.00%) for the first presentation, 90.40% (*SD* = 6.90%) for the second presentation, 91.65% (*SD* = 3.83%) for the third presentation, and 92.25% (*SD* = 4.23%) for the fourth presentation. This ANOVA was significant. Follow-up contrasts indicated that intelligibility scores on the first presentation were worse than intelligibility scores on the second, the third, and the fourth presentations.

Results for Speaker 3 showed a somewhat different pattern from Speakers 1 and 2. Mean intelligibility scores, regardless of presentation mode were as follows: 41.00% (*SD* = 14.92%) for the first presentation, 46.78% (*SD* = 17.01%) for the second presentation, 49.34% (*SD* = 13.72%) for the third presentation, and 54.69% (*SD* = 10.01%) for the fourth presentation. This ANOVA was significant. Follow-up contrasts indicated that intelligibility scores on the first and fourth presentations were significantly different.

Results for Speaker 4 showed the following mean intelligibility scores: 36.30% (*SD* = 7.81%) for the first presentation, 39.15% (*SD* = 12.91%) for the second presentation, 47.80% (*SD* = 12.75%) for the third presentation, and 49.60% (*SD* = 16.71%) for the fourth presentation. This ANOVA was significant. Follow-up contrasts indicated that intelligibility scores on the first presentation were different from intelligibility scores on the third and the fourth presentations.

Results for Speaker 5 showed the same pattern as results for Speaker 3. Mean intelligibility scores, regardless of presentation mode, were as follows: 30.16% (*SD* = 13.83%) for the first presentation, 35.86% (*SD* = 11.12%) for the second presentation, 39.52% (*SD* = 9.83%) for the third presentation, and 42.33% (*SD* = 7.91%) for the fourth presentation. This ANOVA was significant. Follow-up contrasts indicated that intelligibility scores on the first and fourth presentations were significantly different. Individual speaker data are displayed graphically in Figure 3.

Discussion

The present study examined the effects of presentation mode and sequential familiarization on speech intelligibility of 5 individuals with dysarthria secondary to CP. Two individuals with mild dysarthria and 3 individuals with severe dysarthria served as speakers for this study.

Effects of Visual Information on Intelligibility

Across all speakers, results showed that there was a small but significant mean increase in intelligibility when listeners were presented with audiovisual information relative to audio-only information. Although statistically significant, the magnitude of this difference (4%) is probably not clinically meaningful. Further, examination of individual speaker data showed that this overall difference was due to only 1 of the 5 speakers for whom there was a 10% benefit when listeners were given audiovisual input. This individual, Speaker 5, was the most severely impaired with respect to motor control, evidencing more adventitious movement than any other speaker, including involuntary movements of the tongue, forehead, and mandible during speech production. In spite of these severe manifestations of athetosis, listeners were seemingly able to separate meaningful from nonmeaningful facial cues and successfully use the speech-related cues to their benefit for decoding spoken messages.

Results for the other 2 speakers with severe dysarthria,

TABLE 4. Follow-up contrasts for familiarization trials.

Contrast	Mean difference	*df*	Standard error for contrast	*t*
Linear trend (overall)	36.053	1	4.914	7.336*
Trial 1 vs. 2 (overall)	5.695	1	1.216	4.683*
Trial 1 vs. 3 (overall)	8.760	1	1.438	6.092*
Trial 1 vs. 4 (overall)	10.996	1	1.448	7.594*
Trial 2 vs. 3 (overall)	3.066	1	1.341	2.286
Trial 2 vs. 4 (overall)	5.301	1	1.357	3.906*
Trial 3 vs. 4 (overall)	2.236	1	1.127	1.984
Trial 1 vs. 2 (Speaker 1)	8.945	1	2.466	3.627**
Trial 1 vs. 3 (Speaker 1)	8.150	1	2.676	3.045***
Trial 1 vs. 4 (Speaker 1)	8.770	1	2.608	3.363**
Trial 2 vs. 3 (Speaker 1)	0.795	1	1.522	0.522
Trial 2 vs. 4 (Speaker 1)	0.175	1	1.175	0.149
Trial 3 vs. 4 (Speaker 1)	0.620	1	1.596	0.388
Trial 1 vs. 2 (Speaker 2)	5.200	1	1.604	3.242**
Trial 1 vs. 3 (Speaker 2)	6.450	1	1.761	3.662**
Trial 1 vs. 4 (Speaker 2)	7.050	1	2.112	3.338**
Trial 2 vs. 3 (Speaker 2)	1.250	1	1.698	0.736
Trial 2 vs. 4 (Speaker 2)	1.850	1	1.706	1.804
Trial 3 vs. 4 (Speaker 2)	0.600	1	1.082	0.554
Trial 1 vs. 2 (Speaker 3)	5.782	1	3.781	1.529
Trial 1 vs. 3 (Speaker 3)	8.342	1	4.584	1.819
Trial 1 vs. 4 (Speaker 3)	13.695	1	3.432	3.990*
Trial 2 vs. 3 (Speaker 3)	2.560	1	3.731	0.686
Trial 2 vs. 4 (Speaker 3)	7.913	1	3.802	2.081
Trial 3 vs. 4 (Speaker 3)	5.353	1	3.689	1.451
Trial 1 vs. 2 (Speaker 4)	2.850	1	2.961	0.963
Trial 1 vs. 3 (Speaker 4)	11.500	1	3.125	3.680**
Trial 1 vs. 4 (Speaker 4)	13.300	1	4.340	3.065***
Trial 2 vs. 3 (Speaker 4)	8.650	1	3.592	2.408
Trial 2 vs. 4 (Speaker 4)	10.450	1	3.922	2.664
Trial 3 vs. 4 (Speaker 4)	1.800	1	3.199	0.563
Trial 1 vs. 2 (Speaker 5)	5.690	1	2.415	2.358
Trial 1 vs. 3 (Speaker 5)	9.359	1	3.483	2.687
Trial 1 vs. 4 (Speaker 5)	12.164	1	3.308	3.677**
Trial 2 vs. 3 (Speaker 5)	3.663	1	3.489	1.049
Trial 2 vs. 4 (Speaker 5)	6.468	1	3.131	2.066
Trial 3 vs. 4 (Speaker 5)	2.805	1	2.169	1.293

*$p < .001$, **$p < .005$, ***$p < .008$.

Speakers 3 and 4, showed no difference between audiovisual and audio-only input. Closer qualitative examination of these speakers' facial movement, particularly labial and mandibular movement, during speech production suggests characteristics that were different from those of Speaker 5. Perhaps most striking was that both speakers had significant and predominant spasticity that was clearly visible in the external oral and laryngeal regions. This spasticity seemed to be associated with marked reduction of labial and mandibular movement during speech, which in turn may have made it difficult for listeners to extract useful speech-related cues from the audiovisual signal. Overall, the findings of the present study are similar to those of existing studies that have shown inconsistent results for speakers with severe dysarthria.

Results for Speakers 1 and 2, who had mild dysarthria and less severe motor impairments in general, also indicated that there was no difference in intelligibility scores for audiovisual and audio-only presentation. This is somewhat surprising given the consistent finding in the literature that intelligibility is higher for audiovisual than for audio-only presentation modes for non-motor-impaired speakers. One explanation for this finding is that visible facial cues did not provide any unique information beyond what was already available to listeners within the acoustic signal, resulting in the same intelligibility scores for audiovisual and audio-only conditions for these 2 speakers.

From a clinical perspective, findings of the present study suggest that current methods of intelligibility measurement using the isolated auditory signal seem to provide adequate representations of audiovisual as well as audio-only intelligibility scores for most of the individuals examined. Further investigation is necessary to identify the specific visible movement characteristics that enhance intelligibility for speakers with dysarthria.

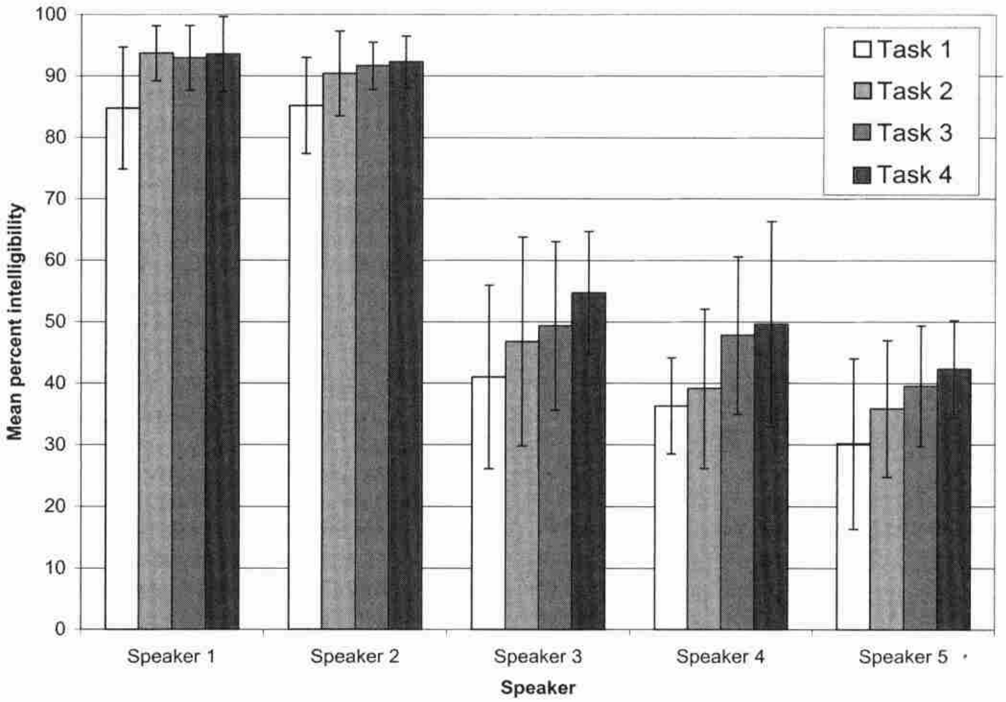
Effects of Sequential Familiarization on Intelligibility

Results of the present study demonstrate that intelligibility scores improve by an average of 11% following four repeated familiarization trials in which unique speech stimuli, spoken by the same individual, are presented to listeners. This finding differs from previous research in two important ways: (a) listeners did not receive feedback or instruction, and (b) listeners heard speakers during four sequential trials over a total of 50 min.

An additional finding was that when listeners' initial familiarization experience involved the audio-only signal followed by the audiovisual signal, greater intelligibility gains occurred between successive trials than for listeners who first received the audiovisual signal followed by the audio-only signal. This suggests that visual information enhances the familiarization effect, perhaps because the multimodal information present in the audiovisual signal facilitates perceptual tuning in listeners.

The primary consistent finding for individual speaker familiarization data was that the first trial always had lower intelligibility scores than the fourth trial, with the mean gain across speakers ranging between 7% and 14%. However, for speakers with severe dysarthria, the benefit of repeated familiarization appeared to accrue slowly, with significant changes in intelligibility observed only between the first and third or first and fourth trials. For speakers with mild dysarthria, the benefit of repeated familiarization occurred immediately, with significant improvements in intelligibility observed between the first and second trials and no change between subsequent adjacent trials. Overall, however, listeners of speakers

FIGURE 3. Mean intelligibility scores (±1 *SD*) by sequential familiarization trial and speaker.



with severe dysarthria tended to show larger familiarization effects across trials than listeners of speakers with mild dysarthria.

Although the methodology and speaker severity characteristics for the present study were different, the familiarization effects observed across trials is generally consistent with the findings of Tjaden and Liss (1995) and Spitzer and colleagues (2000). In addition, results showing no significant difference between intelligibility scores on the first two trials for speakers with severe dysarthria are also consistent with the findings of Garcia and Cannito (1996). The present study extends Garcia and Cannito's results by demonstrating that gains in intelligibility, in the absence of any type of instruction or feedback, are made on subsequent familiarization trials beyond the first two. Together with existing research, results suggest that familiarization can occur under two different circumstances, following a single structured familiarization experience (Tjaden & Liss, 1995; Spitzer et al., 2000) or following repeated sequential listening experiences.

Findings of the present study have several clinical implications. First, results of standard intelligibility measures, which typically do not use familiarization experiences, probably adequately represent the intelligibility experienced by unfamiliar listeners who are one-time communication partners (i.e., wait-staff in restaurants, bank tellers, telemarketers). However, these measures may underestimate the intelligibility scores of frequent or regular communication partners who have repeated experience with the individual's speech. Second, structured familiarization experiences for frequent communication partners may be worthwhile as part of intervention to optimize intelligibility, particularly for individuals with severe dysarthria.

Limitations and Future Directions

There are several limitations to the findings of the present study. Only 5 speakers with dysarthria secondary to CP were included, thus limiting generalizability of findings. Speakers in this study produced a standard corpus of stimulus material that was devoid of communicative intent. As a result, other facial cues and gestures that signal emotion and emphasis were not present. Research exploring the role of this type of information on intelligibility is necessary to understand more fully the effects of visual information on intelligibility in real communication situations.

In the present study, descriptions of movement characteristics of speakers were used to explain some experimental findings. These descriptions were purely qualitative and, consequently, subjective in nature. Future research that provides quantitative measures of oral articulatory and other facial movements is necessary to substantiate the impact of visible oral–facial movement on intelligibility of dysarthric speech.

In each of the four trials in the present study, listeners heard unique speech stimuli produced by the same speaker. It is possible that the intelligibility improvements across trials observed in this study were partly due to listener familiarization with the experimental paradigm as well as familiarization with the dysarthric speech. Alternative research paradigms should be used in future research to determine the differential effects of task familiarization and speaker-specific familiarization.

Finally, future research is necessary to determine the point at which increases in intelligibility associated with listener familiarization reach asymptote, as well as the impact of different familiarization paradigms. The answers to these questions have important clinical implications for enhancing intelligibility of speakers with chronic dysarthria by familiarizing listeners with their speech.

Acknowledgments

We wish to express special thanks to the speakers with dysarthria who participated in this study. We also wish to thank Dr. Robert Prosek for the use of his sound suite to conduct this research. Finally, we acknowledge Amanda Creager, Francesca DiNatale, and Aimee Xander for their assistance with making stimulus tapes and collecting data from listeners.

References

Connolly, J. H. (1986). Intelligibility: A linguistic view. *British Journal of Disorders of Communication, 21*, 371–376.

Erber, N. E. (1975). Auditory–visual perception of speech. *Journal of Speech and Hearing Disorders, 40*, 481–492.

Garcia, J., & Cannito, M. (1996). Influence of verbal and nonverbal contexts on the sentence intelligibility of a speaker with dysarthria. *Journal of Speech, Language, and Hearing Research, 39*, 750–760.

Garcia, J. M., & Dagenais, P. A. (1998). Dysarthric sentence intelligibility: Contribution of iconic gestures and message predictiveness. *Journal of Speech, Language, and Hearing Research, 41*, 1282–1293.

Hoover, J., Reichle, J., Van Tasell, D., & Cole, D. (1987). The intelligibility of synthesized speech: Echo II versus Votrax. *Journal of Speech and Hearing Research, 30*, 425–431.

Hubbard, D. J., & Kushner, D. (1980). A comparison of speech intelligibility between esophageal and normal speakers via three modes of presentation. *Journal of Speech and Hearing Research, 23*, 909–916.

Hunter, L., Pring, T., & Martin, S. (1991). The use of strategies to increase speech intelligibility in cerebral palsy: An experimental evaluation. *British Journal of Disorders of Communication, 26*, 163–174.

Hustad, K. C., & Beukelman, D. R. (2001). Effects of linguistic cues and stimulus cohesion on intelligibility of severely dysarthric speech. *Journal of Speech, Language, and Hearing Research, 44*, 497–510.

Hustad, K. C., Beukelman, D. R., & Yorkston, K. M. (1998). Functional outcome assessment in dysarthria. *Seminars in Speech and Language, 19*(3), 291–302.

Kent, R. (1993). Speech intelligibility and communicative competence in children. In A. P. Kaiser & D. B. Gray (Eds.), *Enhancing children's communication: Foundations for intervention* (Vol. 2, pp. 223–239). Baltimore: Paul H. Brookes.

Kent, R., Weismer, G., Kent, J., & Rosenbek, J. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54*, 482–499.

Kirk, R. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Lindblom, B. (1990). On the communication process: Speaker–listener interaction and the development of speech. *Augmentative and Alternative Communication, 6*, 220–230.

Marascuilo, L. A., & Serlin, R. C. (1988). *Statistical methods for the social and behavioral sciences*. New York: W. H. Freeman.

McNaughton, D., Fallon, K., Tod, F., Weiner, F., & Neisworth, J. (1994). Effect of repeated listening experiences on the intelligibility of synthesized speech. *Augmentative and Alternative Communication, 10*, 161–168.

Monsen, R. B. (1983). The oral speech intelligibility of hearing-impaired talkers. *Journal of Speech and Hearing Disorders, 43*, 286–296.

Munhall, K. G., & Vatikiotis-Bateson, E. (1998). The moving face during speech communication. In R. Campbell, B. Dodd, & D. Burnham (Eds), *Hearing by eye: II. Advances in the psychology of speechreading and auditory-visual speech* (pp. 123–139). Hove, U.K: Psychology Press/Erlbaum.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and noise. *Journal of the Acoustical Society of America, 95*, 1085–1099.

Preminger, J. E., Lin, H. E., Payen, M., & Levitt, H. (1998). Selective visual masking in speechreading. *Journal of Speech, Language, and Hearing Research, 41*, 564–575.

Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *Journal of the Acoustical Society of America, 106*, 3629–3636.

Sanders, D. A., & Goodrich, S. J. (1971). The relative contribution of visual and auditory components of speech to speech intelligibility as a function of three conditions of frequency distortion. *Journal of Speech and Hearing Research, 14*, 154–159.

Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors, 27*, 395–408.

Sonic Foundry. (1998). Sound Forge 4.5 [Computer software]. Madison, WI: Author.

Spitzer, S. M., Liss, J. M., Caviness, J. N., & Adler, C. (2000). An exploration of familiarization effects in the perception of hypokinetic and ataxic dysarthric speech. *Journal of Medical Speech Language Pathology, 8*, 285–293.

Tjaden, K. K., & Liss, J. M. (1995). The role of listener familiarity in the perception of dysarthric speech. *Clinical Linguistics and Phonetics, 9*(2), 139–154.

Venkatagiri, H. (1994). Effect of sentence length and exposure on the intelligibility of synthesized speech. *Augmentative and Alternative Communication, 10*, 96–104.

Yorkston, K., & Beukelman, D. (1980). A clinician-judged technique for quantifying dysarthric speech based on single-word intelligibility. *Journal of Communication Disorders, 13*, 15–31.

Yorkston, K. M., & Beukelman, D. R. (1981). *Assessment of Intelligibility of Dysarthric Speech*. Austin, TX: Pro-Ed.

Yorkston, K., & Beukelman, D. (1983). The influence of judge familiarization with the speaker on dysarthric speech intelligibility. In W. Berry (Ed.), *Clinical dysarthria* (pp. 155–163). San Diego, CA: College Hill.

Yorkston, K. M., Beukelman, D. R., Strand, E. A., & Bell, K. R. (1999). *Management of motor speech disorders in children and adults* (2nd ed.). Austin, TX: Pro-Ed.

Yorkston, K., Beukelman, D., & Tice, R. (1996). Sentence Intelligibility Test for Macintosh [Computer software]. Lincoln, NE: Tice Technology Services.

Yorkston, K., Strand, E., & Kennedy, M. (1996). Comprehensibility of dysarthric speech: Implications for assessment and treatment planning. *American Journal of Speech-Language Pathology, 5*(1), 55–66.

Received June 8, 2002
Accepted July 23, 2002
DOI: 10.1044/1058-0360(2003/066)

Contact author: Katherine C. Hustad, PhD, Department of Communication Sciences & Disorders, The Pennsylvania State University, 110 Moore Building, University Park, PA 16802. E-mail: kch2@psu.edu

Appendix

Instructions for Trials

Instructions for the auditory-only trials were as follows:

You're going to hear a person with a disability producing two sets of 10 sentences. Your job will be to listen carefully and write down what you think the speaker said. You will hear each set of 10 sentences twice—please just listen the first time through. The second time through, there will be time for you to write down what you thought the person said. You can take as much time as you need to do this. This person has significant speech problems, so he or she may be hard to understand. Remember, everything he/she says is a meaningful sentence. Take your best guess and write down anything and everything you thought you understood. We learn as much from what you get wrong as from what you get right.

Instructions for the audiovisual trials were as follows:

You're going to see and hear a person with a disability producing two sets of 10 sentences. Your job will be to listen carefully and write down what you think the speaker said. You will hear each set of 10 sentences twice—please just listen the first time through. The second time through there will be time for you to write down what you thought the person said. You can take as much time as you need to do this. Please watch the person carefully at all times. The purpose of this study is to see whether visual information helps or hinders your ability to understand. This person has significant speech problems, so he or she may be hard to understand. Remember, everything he/she says is a meaningful sentence. Take your best guess and write down anything and everything you thought you understood. We learn as much from what you get wrong as from what you get right.