Original Paper

Folia Phoniatrica et Logopaedica

Folia Phoniatr Logop 2007;59:306–317
DOI: 10.1159/000108337

Effects of Speech Stimuli and Dysarthria Severity on Intelligibility Scores and Listener Confidence Ratings for Speakers with Cerebral Palsy

Katherine C. Hustad

Department of Communicative Disorders and Waisman Center, University of Wisconsin, Madison, Wisc., USA

Key Words
Speech intelligibility · Cerebral palsy · Dysarthria

Abstract
This study examined differences among transcription intelligibility scores and listener confidence ratings for three different types of speech stimuli – single words, unrelated sentences, and sentences forming a narrative – all produced by speakers with dysarthria. Twelve speakers with dysarthria of varying severity secondary to cerebral palsy and 144 listeners participated in this study. Results showed that both intelligibility scores and confidence ratings were differentially affected by both stimuli and severity. For speakers with mild, moderate, and severe dysarthria, intelligibility scores were higher for narratives than for either of the other two types of speech stimuli. For speakers with mild dysarthria, sentences were substantially more intelligible than single words. However, for speakers with moderate, severe, and profound dysarthria, the difference in intelligibility scores for sentences and single words was small or nonsignificant. Confidence ratings did not follow the same pattern as intelligibility data, suggesting a mismatch between listeners' perception of their performance and their actual performance on intelligibility tasks.

Portions of these data were presented at the American Speech-Language and Hearing Association Annual Convention, Chicago, Ill. (November, 2003).

Introduction

Communication occurs at a variety of linguistic levels. In some situations, speakers produce messages comprised of only a single word (e.g. yes, no, hello). In other situations speakers produce utterances comprised of multiple words that form a single thought or idea (i.e. a single sentence). In yet other situations, speakers produce narratives comprised of multiple sentences or ideas that are related in topic. Experimental research suggests that speech intelligibility differs based upon the linguistic level of the spoken message. However, studies have not comprehensively investigated differences among word, sentence, and narrative intelligibility for speakers with dysarthria, and particularly for those who vary in severity. An understanding of how the linguistic level of the spoken message affects intelligibility may have important intervention implications for individuals with dysarthria, helping them to capitalize on the benefits that context may afford.

Word vs. Sentence Intelligibility

Speech intelligibility research has clearly shown that stimuli of varying lengths can result in different intelligibility scores, even for the same speaker. Studies consistently suggest that sentence context tends to yield higher intelligibility scores than words in isolation when speech is mildly to moderately degraded or dysarthric [1–4]. One reason for this phenomenon may be that top-down linguistic-contextual information listeners are able to use different sources of to infer the content of messages may otherwise be degraded beyond recognition.

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1021–7762/07/0596–0306$23.50/0

Accessible online at:
www.karger.com/fpl

Katherine Hustad
475 Waisman Center
1500 Highland Avenue
Madison, WI 53705 (USA)
Tel. +1 608 265 9977, Fax +1 608 263 7710, E-Mail kchustad@wisc.edu

When speech is more severely degraded or dysarthric, the difference between intelligibility of words in isolation and words in sentences seems to become less clear. For example, Miller et al. [1] found that as intelligibility decreased, the difference between intelligibility of words in isolation and words in sentences decreased, although not in a perfectly linear fashion. Their data suggest that when intelligibility was between 15 and 30%, there was no longer a benefit to words in sentences over words in isolation. Research on speakers with hearing impairment supports this conclusion [5]. Studies examining speakers with severe dysarthria are less clear, indicating that individual speaker characteristics may play an important role in determining the difference between sentence and word intelligibility. For some speakers with intelligibility below 30%, sentences continue to be more intelligible than words; for others, words and sentences have similar intelligibility [3, 6], and for others, words may be more intelligible than sentences [4]. The interaction between severity and intelligibility of different stimuli does not appear to be simple or universal.

Word vs. Narrative Intelligibility

Although there have been several studies that have examined word and sentence level intelligibility, few studies have examined narrative intelligibility; thus, little is known about how narrative intelligibility may differ from sentence and word intelligibility. Research on narratives has tended to focus on comprehension [7, 8], which differs from intelligibility in important methodological ways, namely that listeners respond to what they hear in some meaningful way (e.g. answer a question, validate a statement). In contrast, intelligibility studies typically require listeners to report back verbatim what they heard. To make direct comparisons between types of stimuli, it is important that the same dependent measure be used.

Research comparing single-word intelligibility with narrative intelligibility has shown that narrative intelligibility tends to be higher, at least descriptively, than single-word intelligibility for speakers with mild dysarthria [9]. However, word-level intelligibility tends to be higher than or equal to narrative intelligibility for speakers with severe-to-profound dysarthria. For speakers with moderate dysarthria, this difference has not been studied.

Narrative vs. Sentence Intelligibility

Studies comparing narrative intelligibility with sentence intelligibility generally have revealed inconsistent results. For example, Drager and Reichle [10] found that intelligibility of synthesized speech (DECTalk) was significantly higher for sentences presented in a narrative context than for sentences presented in isolation. The magnitude of this difference was approximately 5%. However, in a study examining intelligibility of 4 speakers with severe-profound dysarthria (below 20% intelligibility), Hustad and Beukelman [11] found that intelligibility scores for unrelated sentences and narratives did not differ significantly. One explanation for these discrepant findings relates to the extent of degradation of the speech signal. Drager and Reichle's [10] synthesized speech stimuli were only mildly degraded, whereas Hustad and Beukelman's [11] dysarthric speech stimuli were severely degraded. Thus, it is possible that any advantages that narrative context might provide could be lost when the speech signal degrades to a certain level. Studies have not examined the difference between narrative and sentence intelligibility in moderately degraded speech.

Listeners' Perception of Their Performance

Another factor that may relate to performance on transcription intelligibility tasks is self-efficacy, or listeners' belief that they can be or are successful in deciphering dysarthric speech. Although literature in this area is limited, one preliminary study suggests that listeners tend to underestimate their performance, particularly for speakers with more severe intelligibility deficits [12]. That is, listeners believe that they understand less than they actually do. This mismatch appears to be less pronounced for listeners of speakers with mild dysarthria. However, the relation between performance and perception is unclear for listeners of speakers with moderate dysarthria. Clearly, altered perception of performance may have negative consequences for communicative interchanges between speakers with dysarthria and their partners.

The purpose of the present study was to determine whether there were differences between isolated words, unrelated sentences, and cohesive narratives for speakers with dysarthria of four different severity levels (mild, moderate, severe, profound). Dependent measures of interest were transcription intelligibility scores and also Likert-type ratings of listeners' confidence in their performance. Together, these two measures allowed examination of quantitative and qualitative aspects of performance, providing insight into whether listeners perceived any benefits that were observed via intelligibility scores. The following research questions were addressed: (1) Are there differences among intelligibility scores for the three types of stimuli, words, unrelated sentences, and narratives? Is the same pattern of results evident for speakers

Table 1. Characteristics of speakers with cerebral palsy and dysarthria

Speaker	Age years	Gender	Dysarthria diagnosis	Dysarthria severity	Narrative intelligibility score, %
1	37	M	Spastic	profound	5
2	33	M	Hyperkinetic-spastic	profound	15
3	24	F	Hyperkinetic-spastic	profound	16
4	58	F	Spastic	severe	30
5	46	F	Spastic	severe	37
6	42	F	Spastic	severe	39
7	21	M	Hyperkinetic-spastic	moderate	58
8	33	F	Spastic	moderate	59
9	55	M	Spastic	moderate	65
10	37	M	Spastic	mild	80
11	32	F	Spastic	mild	86
12	53	F	Hyperkinetic-spastic	mild	86

with dysarthria of varying severity? (2) Are there differences among listener confidence ratings for the three types of stimuli? Is the same pattern of results evident for speakers with dysarthria of varying severity?

Method

Participants

Two groups of participants were involved in this research, speakers with dysarthria and nondisabled listeners. Speakers produced stimulus material which was then played for listeners who made orthographic transcriptions of what they heard and made ratings of their confidence in what they had transcribed for each utterance.

Twelve individuals with cerebral palsy, representing a range of severity levels, participated as speakers. Three speakers with mild, moderate, severe, and profound dysarthria, as indicated by intelligibility scores, comprised each severity group. Severity groupings were made retrospectively on the basis of narrative intelligibility scores because narrative data represented the 'best-case scenario' for intelligibility and were thought to represent the functional communication capabilities of each speaker. Speakers with narrative transcription intelligibility scores between 75 and 95% were deemed mild, between 50 and 70% were deemed moderate, between 25 and 45% were deemed severe, and between 0 and 20% were deemed profound. Demographic information for the speakers including age, gender, dysarthria diagnosis, severity, and narrative intelligibility score is provided in table 1. Inclusion criteria required that speakers: (a) use American English as their first and primary language; (b) have normal hearing per self-report; (c) have transcription intelligibility scores for narrative stimuli [11] between 5 and 90%; (d) be between 21 and 55 years of age; (e) be able to produce connected speech consisting of at least 15 consecutive words, and (f) be able to repeat sentences of up to 15 words in length following a verbal model.

One hundred and forty-four individuals without disability participated as listeners. Twelve different individuals were randomly assigned to listen to speech stimuli for each of the 12 speakers with dysarthria. Inclusion criteria required that listeners: (a) use American English as their first and primary language; (b) pass a pure-tone hearing screening at 20 dB SPL for 250 Hz, 500 Hz, 1 kHz, 4 kHz, and 6 kHz bilaterally; (c) have no more than incidental experience listening to or communicating with persons having communication disorders; (d) be between 18 and 40 years of age, and (e) have no identified language, learning, or cognitive disabilities per self-report. Listeners had a mean age of 21.25 years (SD = 2.43). Gender was not a variable of interest; therefore no attempt was made to balance the number of male and female participants.

Materials

Three types of speech stimuli were employed for this study: single words, standard sentences, and pre-scripted narratives. Single-word stimuli were 70 words taken from Kent et al. [13]. These word stimuli were designed for use in phonetic intelligibility testing, with the ultimate aim to provide an explanatory basis for intelligibility deficits in dysarthria. Word stimuli were primarily monosyllabic in nature with CVC or VC phonetic structures. The Kent et al. [13] words have been used extensively in other intelligibility research.

Sentence stimuli were taken from the Sentence Intelligibility Test (SIT) [14]. The SIT is a clinical tool that is widely used to characterize intelligibility of speakers with dysarthria. The SIT consists of a large database of sentences that range in length from 5 to 15 words. Sentences vary in their semantic, syntactic, phonetic, and syllabic composition, with none of these variables controlled or described. The SIT software randomly generates test lists that contain one sentence of each length, for a total of 11 sentences (110 words). Four different randomly generated sentence lists were used in this study.

Narrative stimuli were taken from Hustad and Beukelman [11]. Three different narrative passages, each comprised of 10 related sentences, were used in this study. Passages were developed to represent common situations (e.g. sporting event, natural disaster, purchasing a vehicle), and followed standard American English conventions for content, form, and use of language. Each narrative passage contained a total of 65 words, with the 10 constituent sentences systematically ranging in length from 5 to 8 words. These stimuli have been used in several other research

projects focused on intelligibility of dysarthric speech [11, 15, 16]. The interested reader is referred to Hustad and Beukelman [15] for additional details regarding characteristics of the passages.

Procedures

Recordings of each speaker were made in a quiet environment, either in the speaker's home or in the laboratory. Speakers were recorded on digital audiotape (48 kHz sampling rate; 16-bit quantization) via a professional-quality unidirectional head-mounted microphone positioned 5 cm from the mouth. Speakers produced all stimuli following a model produced by the experimenter. For the narrative stimuli, they were asked to repeat each constituent sentence sequentially. Orthographic representations of stimuli were also provided on a laptop computer situated immediately in front of the speaker. Speakers were required to produce all words verbatim. They were asked to repeat any stimulus sentence that did not include all words per the experimenter's perceptual judgment. Fewer than 5% of stimuli required repetition. Speakers were encouraged to produce stimuli naturally, as they would in real communication situations.

Recorded samples were transferred onto personal computer via a digital sound card, maintaining the sampling rate and quantization of the original recordings. For each speaker, recordings of each stimulus item (individual words from the Kent et al. [13] list, SIT sentences, sentences comprising the narrative passages) were separated into individual sound files. Stimulus files for each sentence were normalized using Sound Forge 4.0 so that the peak amplitude of each stimulus file was constant across all files.

Presentation of each of the three stimulus types was blocked so that listeners heard all single words sequentially, all SIT sentences sequentially, and all sentences comprising the narrative sequentially. The presentation sequence of the 70 word stimuli was randomized for each listener, and the presentation sequence of the 11 SIT sentences was randomized for each listener. Although the narrative stimuli were not randomized, three different narratives were employed and were counterbalanced so that each was represented the same number of times within the speaker groups. It is important to note that narratives were presented in a sentence-by-sentence fashion, identical to presentation of the SIT sentences and single-word stimuli. To prevent a learning effect, presentation of stimulus types (words, sentences, narratives) was counterbalanced, with each of the six possible permutations of presentation order represented two times within each speaker group.

Listeners completed the experiment independently in a soundproof booth. Each listener was seated approximately 2 feet from a high-quality external speaker, with a desktop computer located directly in front of him/her. The presentation level of speech stimuli was calibrated to a peak sound pressure level of 70 dB. Calibration of presentation level was checked periodically to ensure consistency among listeners.

Listeners completed three experimental tasks in which they heard the same speaker producing different types of stimuli (words, SIT sentences, narratives). They were able to hear each stimulus item only one time. Experimental tasks were presented via computer using a custom-programmed setup in Microsoft Powerpoint. Prior to beginning the experimental tasks, listeners were instructed that they would hear a person with a speech problem who would be producing words and sentences. They were instructed to type exactly what they thought the speaker said between each sentence for the sentence and narrative stimuli and between each word for the word stimuli. Listeners were also informed that the person speaking would be difficult to understand and that if they were uncertain of what the speaker said, they should venture their best guess. For each of the three intelligibility tasks, stimulus items were presented one sentence at a time (for the sentence and narrative tasks) or one word at a time (for the word task). After presentation of individual items, listeners were prompted to type what they heard. After transcribing each item, listeners responded to the question: 'How confident are you that what you typed is correct?' To do this, they were presented with a visual display of a Likert scale with two anchor points (1 = not confident at all; 7 = extremely confident) and asked to type a number into a text box. Listeners then advanced to the next stimulus item. After advancing, they were unable to return to previous items or to edit previous responses. Listeners were instructed that they could take as much time as necessary and that the experiment was completely self-paced.

Scoring and Reliability

Intelligibility scores were obtained by calculating the percent of words transcribed correctly for each experimental task and listener. Transcriptions from each listener were scored by the experimenter, who tallied the number of words identified correctly based upon whether each was an exact phonemic match to the corresponding word in the target utterance. In this paradigm, individual words were considered correct if the phonemes represented by the orthographic transcription were consistent with the phonemes in the target stimulus item (e.g. they're = their; two = too; feet = feat). For the word-length stimuli, there were 70 possible words. For the SIT sentences, there were 110 possible words across the 11 sentences. For the narrative passages, there were 65 words across the 10 constituent sentences. Because the stimuli had different denominators, the number of words transcribed correctly was converted to a percent score to allow comparison among stimuli.

Listener confidence ratings were the values (ranging from 1 to 7; 1 = not at all confident; 7 = very confident) that listeners entered into the computer following transcription of each utterance. Thus, for each utterance transcribed, there was an analogous confidence rating.

Interlistener reliability was determined by examining agreement in transcription of individual words for 2 randomly selected listeners from each speaker (16.67% of the sample). To do this, one sentence from the narrative stimuli, one sentence from the SIT stimuli, and 3 words from the word stimuli were selected by the experimenter. Listener transcripts for the target stimuli were evaluated for word-by-word agreement by tallying the number of occurrences in which both listeners got the same individual words correct, both listeners got the same individual words incorrect, or both listeners omitted the same individual word. Word-by-word disagreement was tallied by counting the number of occurrences in which one listener identified the word correctly and the other listener did not, or one listener omitted the word and the other did not. Word-by-word transcription agreement across all 24 listeners, calculated using the formula

$$\text{percent agreement} = \text{agreements}/(\text{agreements} + \text{disagreements}) \times 100,$$

was 81.94%. Intralistener reliability was not obtained because of documented learning effects that occur within 1 h or less of exposure to dysarthric speech [17].

Table 2. Descriptive data for intelligibility scores (Intell), and confidence ratings (Conf) by individual speaker within each severity group

Speaker	Severity	Descriptives	Words		Sentences		Narratives	
			Intell	Conf	Intell	Conf	Intell	Conf
Speaker 1	profound	mean	5.17	2.58	6.97	1.72	4.23	1.89
		SD	2.15	0.82	4.03	0.80	4.10	0.90
Speaker 2	profound	mean	16.33	4.03	11.52	2.30	14.87	2.70
		SD	4.05	1.12	3.17	0.97	6.61	1.03
Speaker 3	profound	mean	9.92	2.69	11.56	1.16	16.05	1.69
		SD	4.03	0.87	5.59	0.36	7.45	0.61
Speaker 4	severe	mean	13.75	3.24	27.49	2.60	29.24	2.91
		SD	3.17	0.71	4.30	0.86	8.97	1.15
Speaker 5	severe	mean	10.03	3.22	16.36	1.94	37.28	3.01
		SD	3.28	0.60	5.98	0.58	14.87	1.20
Speaker 6	severe	mean	31.26	4.62	29.70	3.14	38.95	3.56
		SD	3.81	1.01	5.40	0.89	13.70	1.56
Speaker 7	moderate	mean	32.58	4.31	30.23	2.69	58.08	3.75
		SD	6.11	0.71	6.79	1.04	11.85	1.38
Speaker 8	moderate	mean	27.50	3.72	28.48	2.77	58.72	4.07
		SD	10.12	1.04	4.86	0.93	14.93	0.79
Speaker 9	moderate	mean	36.38	3.81	39.69	2.65	64.66	4.43
		SD	6.35	0.74	8.61	0.80	9.90	0.97
Speaker 10	mild	mean	51.63	4.62	70.53	4.35	80.26	5.16
		SD	7.76	0.95	11.53	0.85	14.47	1.35
Speaker 11	mild	mean	49.75	4.46	69.95	4.49	85.77	5.86
		SD	9.66	0.90	8.64	0.72	10.17	0.76
Speaker 12	mild	mean	60.67	8.24	59.27	4.58	86.28	5.86
		SD	6.12	0.64	5.33	0.60	12.32	0.83

Experimental Design and Statistical Procedures

Three analyses were employed for this study. For each set of analyses, 3 speakers comprised each severity group (mild, moderate, severe, profound), with 12 different listeners providing transcriptions for each speaker, for a total of 36 different listeners per severity group.

The first analysis examined differences among intelligibility scores for the three types of speech stimuli. A 3 × 4 split-plot design [16] and a fully factorial parametric ANOVA were employed for this analysis. The within-subjects variable was 'Stimuli' and its three categories were words, sentences, and narratives. The between-subjects variable was 'Severity' and its four groups were mild, moderate, severe, and profound. Follow-up contrasts focused on differences in intelligibility scores within severity groups.

The second analysis evaluated differences among listener confidence ratings for each of the three different types of speech stimuli. The same 3 × 4 split-plot design [18] was employed. However, because confidence ratings were ordinal in nature, nonparametric analyses were used. Limitations inherent to nonparametric statistics permitted examination of only the main effect of stimulus type (via the Friedman test). Follow-up contrasts focused on differences in confidence ratings within severity groups and were performed using the Wilcoxon signed-rank test.

The third analysis evaluated relationships between confidence ratings and intelligibility scores for each of the three different types of stimuli. For this analysis, nonparametric correlation coefficients using Kendall's tau were calculated for each of the three types of stimuli within each of the three severity groups.

Results

Differences in Intelligibility among Types of Stimuli

Descriptive statistics are provided in table 2 and are shown graphically in figure 1. Descriptive data suggest that there were no differences in mean intelligibility scores for the three types of stimuli for the profound group. However, for the severe, moderate, and mild groups, narratives tended to have higher intelligibility scores than words and sentences. There was a small descriptive advantage for sentences over words for the severe and mild groups, but not for the moderate group.

ANOVA results showed that the main effect of stimuli was significant ($F = 199.92$; $p < 0.001$). The main effect of severity was also significant ($F = 495.56$; $p < 0.001$).

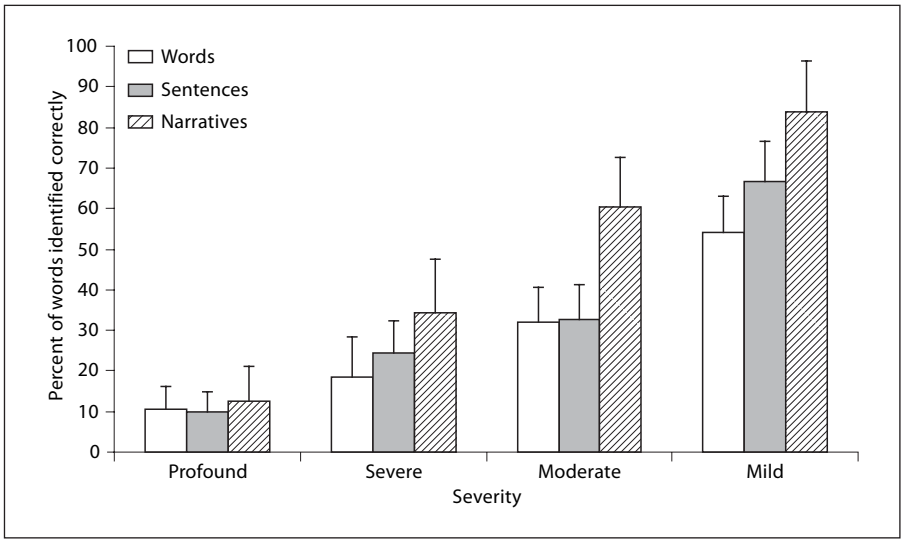

Fig. 1. Transcription intelligibility scores by stimulus type and speaker severity (percent of words transcribed correctly, mean ± SD).

Table 3. Follow-up contrasts comparing intelligibility scores from different stimuli within severity groups

Contrast	Mean difference % correct	d.f.	SE	t
Sentences – narrative (mild)	–17.35	35	2.26	–7.65*
Sentences – words (mild)	12.57	35	2.29	5.48*
Narratives – words (mild)	29.92	35	2.25	13.27*
Sentences – narrative (moderate)	–27.58	35	2.32	–11.88*
Sentences – words (moderate)	0.64	35	1.30	0.49
Narratives – words (moderate)	28.23	35	2.47	11.40*
Sentences – narrative (severe)	–9.97	35	2.32	–4.29*
Sentences – words (severe)	6.17	35	1.54	4.02*
Narratives – words (severe)	16.14	35	2.51	6.42*
Sentences – narrative (profound)	–2.59	35	1.36	–1.90
Sentences – words (profound)	–0.45	35	1.06	–0.43
Narratives – words (profound)	2.13	35	1.34	1.59

Note that in the column labeled 'Contrast' the second variable is subtracted from the first to get the mean difference.

* $p < 0.004$ (0.05/12 contrasts).

Finally, the interaction between stimuli and severity was significant ($F = 26.52$; $p < 0.001$).

Follow-up contrasts focused on intelligibility differences within severity groups and are presented in table 3. For speakers with mild dysarthria, narrative stimuli resulted in significantly higher intelligibility scores than sentence stimuli or word stimuli; and sentence stimuli resulted in higher intelligibility scores than word stimuli. For speakers with moderate dysarthria, again narrative stimuli resulted in significantly higher intelligibility scores than sentence stimuli or word stimuli; however, the difference between sentence stimuli and word stimuli was not significant. Speakers with severe dysarthria showed the same results as those with mild dysarthria: narrative stimuli resulted in significantly higher intelligibility scores than sentence stimuli or word stimuli; and sentence stimuli resulted in higher intelligibility scores than word stimuli. For speakers with profound dysarthria, there were no significant differences in intelligibility scores for any of the three types of stimuli.

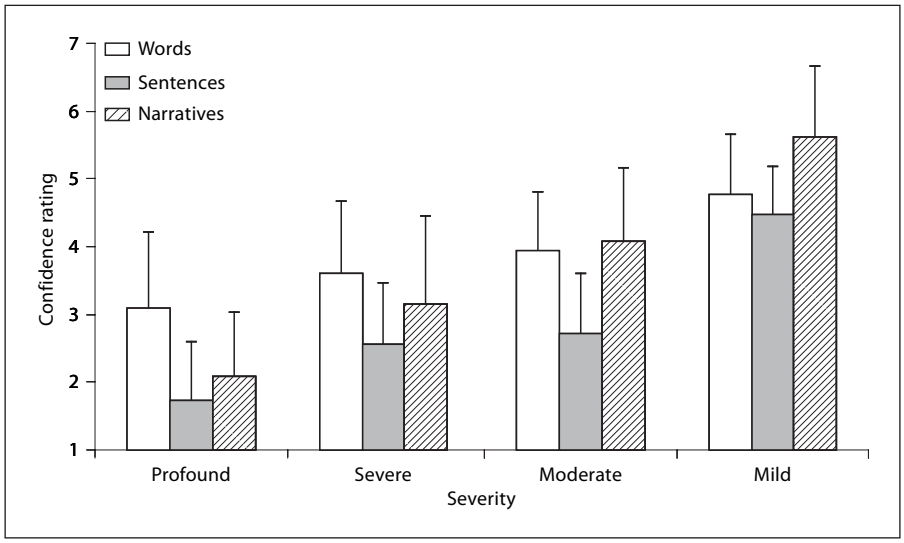

Fig. 2. Confidence ratings for intelligibility transcriptions by stimulus type and speaker severity (mean ± SD). Note that 1 = not at all confident; 7 = very confident.

Table 4. Nonparametric analyses comparing confidence ratings for intelligibility transcriptions for different stimuli within severity groups using the Wilcoxon signed-rank test (large sample approximation)

Contrast	Mean difference Likert points	Ranks				
		negative	positive	ties	n	Z
Sentences – narratives (mild)	–1.16	5	31	0	36	–4.76*
Sentences – words (mild)	–0.30	12	24	0	36	–2.52
Narratives – words (mild)	0.86	30	6	0	36	–4.08*
Sentences – narratives (moderate)	–1.37	4	32	0	36	–4.85*
Sentences – words (moderate)	–1.23	4	32	0	36	–4.61*
Narratives – words (moderate)	0.14	21	15	0	36	–1.19
Sentences – narratives (severe)	–0.61	9	27	0	36	–3.02*
Sentences – words (severe)	–1.13	5	31	0	36	–4.88*
Narratives – words (severe)	–0.52	12	24	0	36	–2.40
Sentences – narratives (profound)	–0.36	8	23	5	36	–3.25*
Sentences – words (profound)	–1.36	1	35	0	36	–5.22*
Narratives – words (profound)	–1.00	2	34	0	36	–4.90*

Note that in the column labeled 'Contrast' the second variable is subtracted from the first to get the mean difference.

* $p < 0.004$ (0.05/12 contrasts).

Differences in Confidence Ratings among Types of Stimuli

Again, descriptive statistics are provided in table 2 and are shown graphically in figure 2. Descriptive data suggest that listeners in the severe and profound groups were most confident when presented with word stimuli. However, listeners in the mild and moderate groups were most confident when presented with the narrative stimuli. Listeners within all speaker groups were least confident when presented with sentence stimuli.

The nonparametric Friedman's omnibus test showed that the main effect of stimuli was significant ($\chi^2 = 93.34$; $p < 0.001$). Follow-up contrasts focused on differences in confidence ratings within severity groups and are present-

ed in table 4. For speakers with mild dysarthria, narrative stimuli resulted in significantly higher confidence ratings than sentence stimuli or word stimuli, and confidence ratings for sentence stimuli did not differ from those for word stimuli. For speakers with moderate dysarthria, again narrative stimuli resulted in significantly higher confidence ratings than sentence stimuli. In addition, word stimuli resulted in higher confidence ratings than sentence stimuli, and the difference between narrative stimuli and word stimuli was not significant. Speakers with severe dysarthria showed the same results as those with moderate dysarthria: narrative stimuli resulted in significantly higher confidence ratings than sentence stimuli; word stimuli resulted in higher confidence ratings than sentence stimuli, and the difference between narrative stimuli and word stimuli was not significant. For speakers with profound dysarthria, narrative stimuli resulted in significantly higher confidence ratings than sentence stimuli, and word stimuli resulted in higher confidence ratings than sentence stimuli and narrative stimuli.

Relationships between Confidence Ratings and Intelligibility Scores

Overall, relationships between confidence ratings and intelligibility scores, although positive in direction, were weak. Statistics are provided in table 5. Scatterplots illustrating the relationships between confidence ratings and intelligibility scores are shown in figures 3–6. The only correlation coefficients that were statistically significant ($p < 0.001$) were those between confidence ratings and narrative intelligibility scores for speakers in the mild, moderate, and severe groups. However, even these significant correlations were only moderately strong.

Table 5. Nonparametric correlations between confidence ratings and intelligibility scores by stimuli and severity group using Kendall's tau

Severity	Stimuli		
	words	sentences	narratives
Mild	0.306	0.148	0.442*
Moderate	0.243	0.168	0.506*
Severe	0.287	0.336	0.418*
Profound	0.311	0.124	0.068

* $p < 0.001$.

Discussion

This study examined differences among intelligibility scores and listener confidence ratings for three different types of speech stimuli – single words, unrelated sentences, and sentences forming narratives. Stimuli were produced by 12 speakers with dysarthria of varying severity. Results showed that intelligibility scores and confidence ratings were differentially affected by both stimuli and speaker severity. Most noteworthy was that narrative stimuli had a consistent and powerful positive effect on intelligibility for speakers with mild, moderate, and severe dysarthria relative to the other types of stimuli. Sentence stimuli also had a positive effect on intelligibility, relative to single-word stimuli, but only for speakers with mild dysarthria. Confidence ratings did not follow the same pattern as intelligibility data. In addition, correlations between intelligibility scores and confidence ratings, although positive in direction, were generally weak, except for the narrative stimuli, which showed moderate correlations within the mild, moderate, and severe groups. Results suggest that there may be a mismatch between listeners' perception of their performance and their actual performance on intelligibility tasks for some types of stimuli.

Effects of Stimulus Length and Severity on Intelligibility Scores

It was hypothesized that narratives would yield higher intelligibility scores than sentences or words because the related nature of the sentences comprising the narrative would facilitate the building of contextual knowledge in a progressive fashion for each sentence. This building of context would facilitate listeners' ability to use top-down intrinsic linguistic knowledge, allowing them to infer words that they may not otherwise have been able to identify.

Results confirmed this hypothesis for speakers with mild, moderate and severe dysarthria. The magnitude of the difference between narratives and the other types of stimuli ranged from 10 to 30% on average, suggesting that the context associated with narratives provided an important advantage to listeners. Further, this result suggests that the ability and opportunity to invoke top-down linguistic knowledge plays a very important role in helping listeners to decipher dysarthric speech. The positive impact of narrative context on intelligibility scores in the present study is consistent with the findings of Drager and Reichle [10] and suggests that narrative context may have a greater effect on intelligibility of dysarthric speech

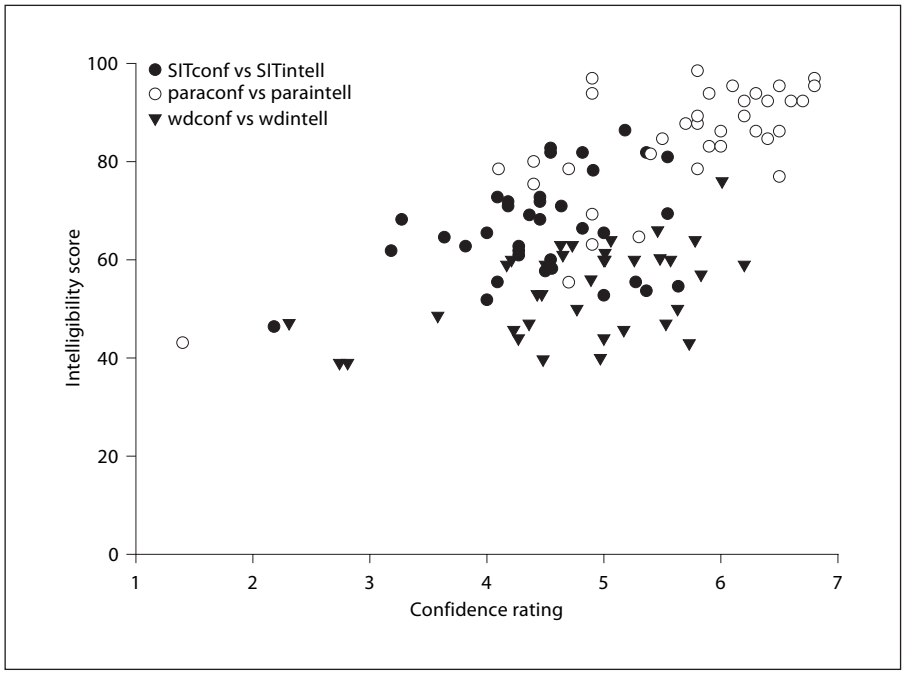

Fig. 3. Intelligibility scores as a function of confidence ratings by individual speaker and stimulus type for the mild dysarthria group. SITconf = Confidence rating for the SIT sentences; SITintell = intelligibility scores for the SIT sentences; paraconf = confidence ratings for the sentences comprising a narrative; paraintell = intelligibility scores for the sentences comprising a narrative; wdconf = confidence ratings for single word stimuli; wdintell = intelligibility scores for single word stimuli.

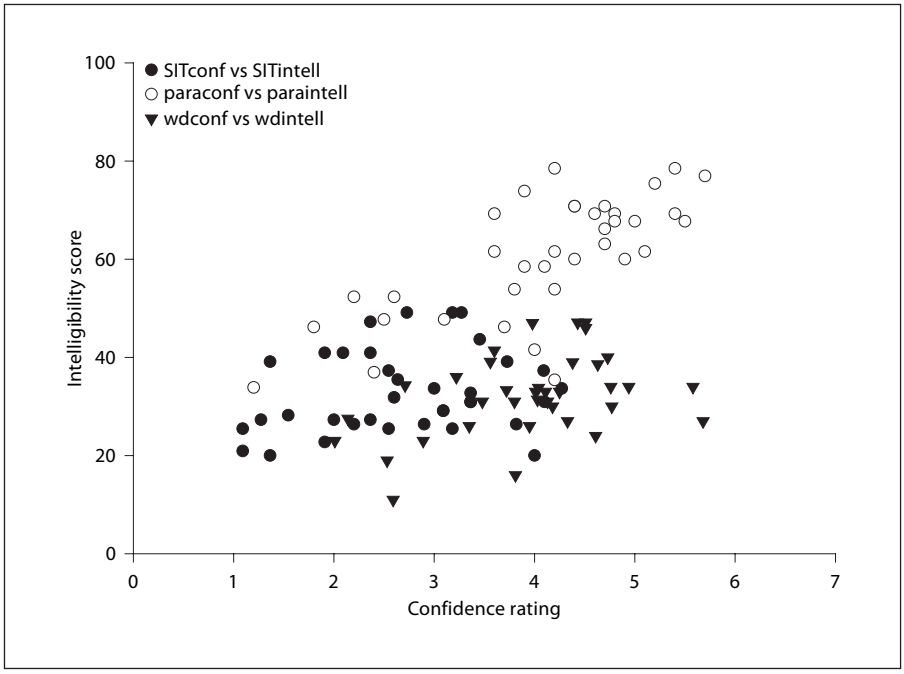

Fig. 4. Intelligibility scores as a function of confidence ratings by individual speaker and stimulus type for the moderate dysarthria group. For abbreviations see legend to figure 3.

than on synthesized speech. Similar to the findings of Hustad and Beukelman [11], the present study showed that for speakers with profound dysarthria, there was no intelligibility advantage for any of the three types of stimuli. One reason may be that listeners were not able to identify enough content information from the speech signal to allow them to make optimal use of top-down linguistic knowledge. Thus, there was no benefit and, interestingly, no detriment to connected speech stimuli over single words.

It was also hypothesized that sentences would yield higher intelligibility scores than single words, again because contextual information from each sentence would facilitate listeners' use of top-down linguistic contextual

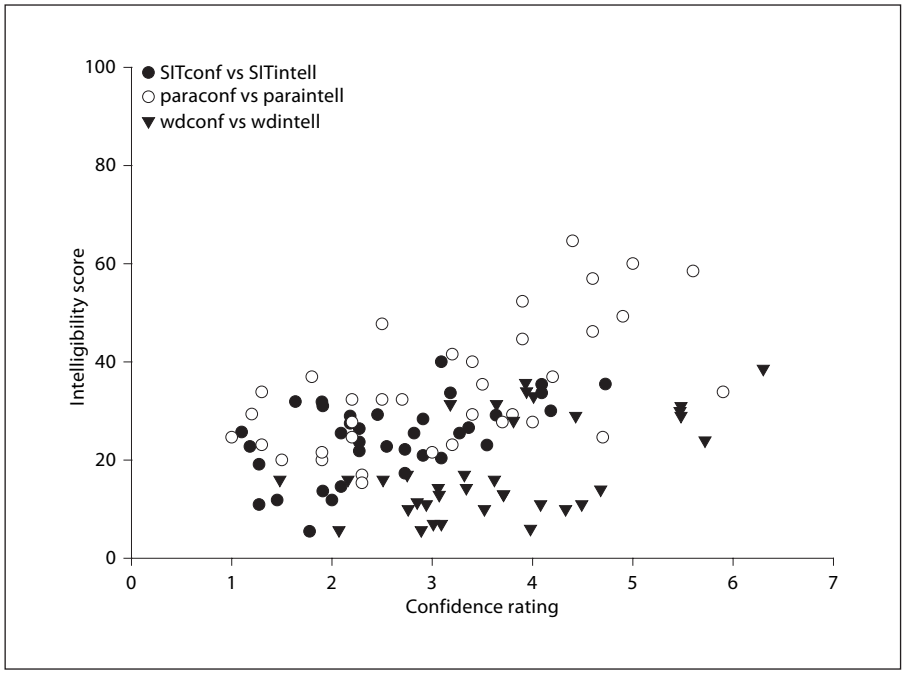

Fig. 5. Intelligibility scores as a function of confidence ratings by individual speaker and stimulus type for the severe dysarthria group. For abbreviations see legend to figure 3.

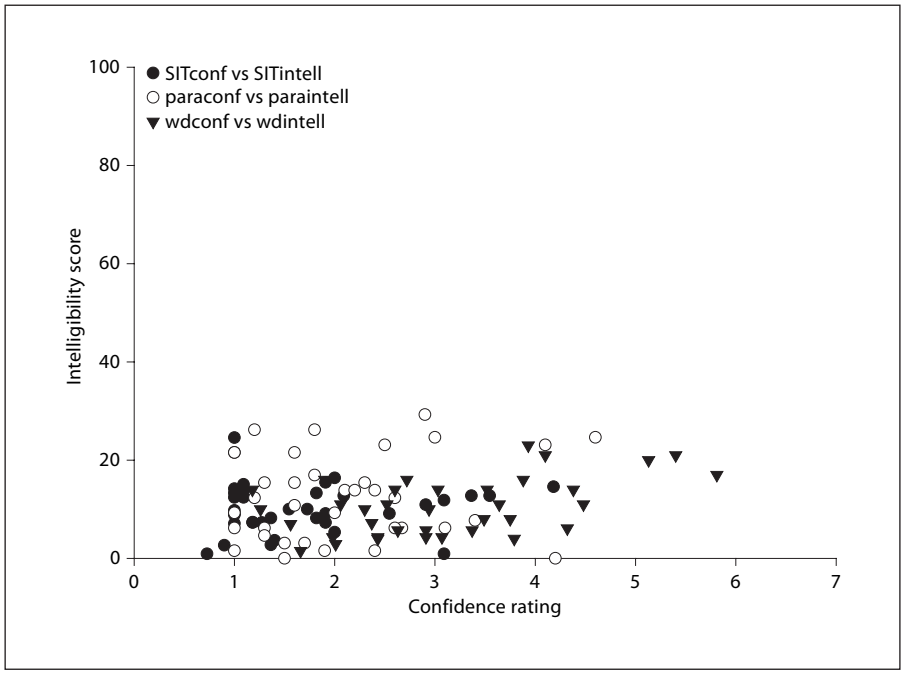

Fig. 6. Intelligibility scores as a function of confidence ratings by individual speaker and stimulus type for the profound dysarthria group. For abbreviations see legend to figure 3.

knowledge for identifying constituent words. Consistent with previous studies [1–4], results of the present study confirmed this hypothesis for speakers with mild dysarthria. However, the difference between word and sentence intelligibility was not significant for speakers with moderate and profound dysarthria, and although statistically significant, it was very small for speakers with severe dysarthria. One explanation for the small or nonexistent difference between word and sentence intelligibility for speakers with moderate to profound dysarthria may relate to the nature of the sentence stimuli from the SIT [14]. As described previously, sentences systematically varied in length. Longer sentences tended to have more complex, and subsequently less predictable, syntactic constructions

(e.g. embedded clauses), which may have made it difficult for listeners to use linguistic knowledge to their advantage in deciphering constituent words. Another explanation relates to production capabilities of the speakers. In particular, speakers with greater intelligibility limitations tend to be those with greater motor control challenges. Therefore, it is likely that these speakers had more difficulty producing longer sentences and consequently the quality of their productions was reduced. Any inherent benefit from sentential context may have been, in effect, canceled by production characteristics that were suboptimal for the more involved speakers. Studies examining specific production changes associated with increasing length and complexity of stimuli are necessary to validate this speculation.

Effects of Stimulus Length and Severity on Confidence Ratings

It was hypothesized that listeners would have reasonable, but probably not perfect, insight into their performance on intelligibility tasks. Thus, confidence ratings would generally mirror the same pattern of results as intelligibility scores. This was true in the most general sense, i.e. confidence ratings tended to decrease as severity increased. However, confidence ratings were differentially affected by both stimuli and severity in a different and somewhat less predictable way than observed for intelligibility scores. For speakers with moderate, severe, and profound dysarthria, confidence ratings were higher for words than for sentences. Intelligibility scores for the same speakers indicated little or no difference in intelligibility scores between these two types of stimuli. One explanation for this finding may be that listeners were more confident in word intelligibility performance because the task was simpler and less complex. In addition, lower confidence ratings for sentences may reflect increased production difficulties on behalf of the speakers.

For speakers with moderate and severe dysarthria, confidence ratings for narratives and words did not differ, even though intelligibility scores were significantly higher for narratives. This finding is difficult to interpret; however, one explanation may, again, be that listener confidence in word transcriptions was inflated relative to their actual performance because transcribing single words was simpler and less taxing than transcribing connected speech. The same explanation may also suffice for the finding that confidence ratings for words were higher than both sentences and narratives for speakers with profound dysarthria.

For speakers with mild dysarthria, narratives resulted in the highest confidence ratings, which is consistent with intelligibility scores and suggests that listeners had relatively accurate insight into their performance. Confidence ratings for words and sentences did not differ for speakers with mild dysarthria, even though intelligibility scores were higher for sentences than for words. This finding may relate to the nature of the sentences employed in this study, perhaps suggesting that listeners perceived their transcriptions to be less accurate than they actually were because some of the sentences were lengthy and linguistically complex in nature. Another possibility is that confidence ratings reflect the increased processing or working memory load placed on listeners when stimuli were more complex and lengthy.

Relationships between Confidence Ratings and Intelligibility Scores

Results of this study showed that the strongest relationships between confidence ratings and intelligibility scores occurred when listeners transcribed narrative stimuli (with the exception of speakers with profound dysarthria). In general all other relationships were weak and nonsignificant, indicating that listener confidence ratings, regardless of speaker severity, were not closely linked to their performance on intelligibility tasks for unrelated sentences and for single words. This finding provides support for the notion that confidence ratings may be a proxy for some other phenomena such as processing load or working memory.

Limitations and Future Directions

The present study was experimental in nature. Speech stimuli were scripted and not spontaneous. The listening environment was carefully controlled and optimized. Listeners could only hear speakers and were not able to see them. As a result, the extent to which findings of the present study might generalize to real-life speaking situations is limited. Additional research is necessary to investigate a myriad of variables related to the influence of speech stimuli on intelligibility. Questions include: (1) how spontaneous narratives might compare with spontaneous sentences and words with regard to intelligibility; (2) how visual information might influence intelligibility of narratives, sentences, and words; (3) how different types of sentences that are controlled for syntactic, semantic, and morphologic features might compare with narrative and word intelligibility, and (4) how the opportunity for speakers and listeners to interact on-line and face-to-face might influence intelligibility of different speech stimuli.

Conclusions

Results of the present study demonstrate that the nature of speech stimuli produced by speakers with dysarthria has an important effect on intelligibility. In particular this study showed that narrative context has an especially powerful effect on intelligibility, particularly for speakers with severe, moderate and mild dysarthria. In addition, correlations between confidence ratings and intelligibility scores were highest for narratives. Presumably, narratives enhance intelligibility because context builds from sentence to sentence, enabling listeners to effectively use top-down linguistic knowledge to aid in inferring words that may be difficult to decode. For speakers with profound dysarthria, no such contextual benefit was observed for intelligibility, suggesting that perhaps there is some minimal number of words that must be decoded correctly in order for listeners to make use of certain types of linguistic knowledge.

The present study also showed that sentence context had a positive impact on intelligibility; however, the magnitude and extent of this effect was less than that of narratives. In particular, speakers with mild dysarthria showed the greatest benefit from sentence context. Speakers with more marked intelligibility deficits showed little or no benefit from sentence context, relative to single words.

Finally, results of this study revealed that listeners' insight into their performance on transcription intelligibility tasks was variable and that in some cases listeners' confidence in their ability to decipher dysarthric speech was inappropriately low, and in other cases, it was inappropriately high, especially when the task appeared to be a simple one (i.e. transcribing single words). Additional research is necessary to investigate how listeners' confidence in their ability to transcribe dysarthric speech might influence the success of interactions with speakers who have dysarthria and the strategies that listeners employ.

Acknowledgments

The author thanks Julie Auker, Katie Rentschler, Gwen Gottardy, Heidi Ake, Becky Lang, and Rhonda Davis for assistance with data collection from listeners. In addition, the author thanks Caitlin Dardis and Lisa Igl for assistance with scoring transcriptions. This research was funded by grant R03 DC005536 from the National Institute on Deafness and Other Communication Disorders (NIH), and by grant P30 HD03352 to the Waisman Center from the National Institute of Child Health and Human Development (NIH).

References

1 Miller GA, Heise GA, Lichten W: The intelligibility of speech as a function of the context of the test materials. J Exp Psychol 1951; 41:329–335.
2 O'Neil JJ: Recognition of intelligibility test materials in context and isolation. J Speech Hear Disord 1957;22:87–90.
3 Yorkston K, Beukelman D: A comparison of techniques for measuring intelligibility of dysarthric speech. J Commun Disord 1978; 11:499–512.
4 Yorkston K, Beukelman D: Communication efficiency of dysarthric speakers as measured by sentence intelligibility and speaking rate. J Speech Hear Disord 1981;46:296–301.
5 Sitler RW, Schiavetti N, Metz DE: Contextual effects in the measurement of hearing-impaired speakers' intelligibility. J Commun Disord 1983;11:22–30.
6 Dongilli P: Semantic context and speech intelligibility; in Till J, Yorkston K, Beukelman D (eds): Motor Speech Disorders: Advance in Assessment and Treatment. Baltimore, Brookes, 1994, pp 175–191.
7 Giolas T, Epstein A: Comparative intelligibility of word lists and continuous discourse. JSHR 1963;6:349–358.
8 Giolas TG: Comparative intelligibility scores of sentence lists and continuous discourse. J Aud Res 1965;6:31–38.
9 Beukelman DR, Yorkston K: The relationship between information transfer and speech intelligibility of dysarthric speakers. J Commun Disord 1979;12:189–196.
10 Drager KDR, Reichle JE: Effects of discourse context on the intelligibility of synthesized speech for young adult and older adult listeners: applications for AAC. J Speech Lang Hear Res 2001;44:1052–1057.
11 Hustad KC, Beukelman DR: Effects of linguistic cues and stimulus cohesion on intelligibility of severely dysarthric speech. J Speech Lang Hear Res 2001;44:497–510.
12 Hustad KC: Estimating the intelligibility of speakers with dysarthria. Folia Phoniatr Logop 2006;58:217–228.
13 Kent R, Weismer G, Kent J, Rosenbek J: Toward phonetic intelligibility testing in dysarthria. J Speech Hear Disord 1989;54:482–499.
14 Yorkston K, Beukelman D, Tice R: Sentence Intelligibility Test for Windows. Lincoln, Communication Disorders Software. Distributed by the Institute for Rehabilitation Science and Engineering, Lincoln, 1996.
15 Hustad KC, Beukelman DR: Listener comprehension of severely dysarthric speech: effects of linguistic cues and stimulus cohesion. J Speech Lang Hear Res 2002;45: 545–558.
16 Hustad KC, Jones T, Daily S: Implementing speech supplementation strategies: effects on intelligibility and speech rate of individuals with chronic severe dysarthria. J Speech Lang Hear Res 2003;46:462–474.
17 Hustad KC, Cahill MA: Effects of presentation mode and repeated familiarization on intellligibility of dysarthric speech. Am J Speech Lang Pathol 2003;12:1–11.
18 Kirk R: Experimental Design: Procedures for the Behavioral Sciences, ed 3. Pacific Grove, Brooks/Cole, 1995.