

A Closer Look at Transcription Intelligibility for Speakers With Dysarthria: Evaluation of Scoring Paradigms and Linguistic Errors Made by Listeners

Katherine C. Hustad

University of Wisconsin—Madison

Purpose: This study addressed the effects of 3 different paradigms for scoring orthographic transcriptions of dysarthric speech on intelligibility scores. The study also examined whether there were differences in transcription accuracy among words from different linguistic classes.

Method: Speech samples were collected from 12 speakers with dysarthria of varying severity. Twelve different listeners made orthographic transcriptions of each speaker, for a total of 144 listeners. Transcriptions were scored using 3 different paradigms: total word phonemic match, informational word phonemic match, and informational word semantic match. Transcriptions were also coded into 3 linguistic categories: content words, modifiers, and functors. The number of words that each listener transcribed correctly within each category was tallied.

Results: There were significant differences among the 3 scoring paradigms. However, the magnitude of differences was relatively small. In addition, listeners transcribed functor words more accurately than content words or modifiers. They also transcribed free morphemes more accurately than bound morphemes.

Conclusions: The specific scoring paradigm that clinicians employ for measuring speech intelligibility appears to be relatively inconsequential as long as consistent procedures are used. Analyses of the kinds of words that listeners transcribe correctly suggest that interventions focusing on listener processing strategies should be considered for enhancing intelligibility of speakers with chronic dysarthria.

Key Words: speech intelligibility, speech perception, dysarthria, cerebral palsy

Speech intelligibility is a complex and multifaceted construct. Improved intelligibility is often a primary goal of speech therapy, especially for individuals with dysarthria (Ansel & Kent, 1992; Yorkston, Beukelman, Strand, & Bell, 1999). However, many aspects of intelligibility and the variables that contribute to it remain poorly understood. Essential to the construct of intelligibility is (a) a speaker who produces an acoustic signal for the purposes of conveying linguistic content and (b) a listener who receives the signal and interprets the linguistic content (Yorkston et al., 1999). Thus, both production-related variables associated with the speaker and perception-related variables associated with the listener play key roles in intelligibility.

One long-standing paradigm for measuring intelligibility of speakers with dysarthria has involved orthographic transcription of the acoustic signal by listeners. From listener transcriptions, “percentage intelligibility” can be obtained by dividing the number of words identified correctly by the

number of words possible (Tikofsky & Tikofsky, 1964; Yorkston & Beukelman, 1978, 1980; Yorkston et al., 1999). Most commonly, transcribed words are counted as “correct” if they match the target word phonemically. However, this criterion has been applied in different ways. For example, some studies have required that transcribed words be an exact phonemic match to target words (see Hustad, Jones, & Dailey, 2003). Other studies have evaluated the match between target and transcribed words without regard for small morphological errors affecting tense and number as long as those errors did not change the syllabic structure of transcribed words (see Liss, Spitzer, Caviness, & Adler, 2002). Although the extent to which these methodological differences influence intelligibility scores is unknown, it seems likely that the former would result in lower intelligibility scores.

Another issue with transcription intelligibility measures relates to the weighting of individual words in the final

intelligibility score. Clearly, some words carry more information than others and consequently make a more important contribution to the overall meaning of the utterance. However, in the standard transcription intelligibility paradigm, each word is weighted equally. Thus, words that carry little meaning (e.g., *a, the, in*) make the same contribution to the intelligibility score as words that carry important content. Monsen (1978, 1983) addressed this issue in his work examining intelligibility of speakers with hearing impairment. In these studies, Monsen weighted individual words on the basis of how much each word contributed to the content expressed in the target sentences. Unfortunately, he did not compare the weighted scores with unweighted scores, so the impact of this type of semantic weighting is unknown. It is unclear whether differences among scoring paradigms such as those described here yield important differences in intelligibility results, but the answer may have clinical implications. Toward this end, the first goal of the present study was to quantify the extent to which differences in scoring paradigms affect resultant intelligibility scores.

Although the intelligibility score is often regarded as an attribute of the speaker, it is also a reflection of what a listener hears when presented with a particular speech signal. Few studies have closely examined listener performance, beyond the basic intelligibility score. One exception is the work of Liss, Spitzer, Caviness, Adler, and Edwards (1998, 2000). In this series of studies, Liss and colleagues examined lexical segmentation errors made by listeners when presented with the speech of individuals with ataxic and hypokinetic dysarthria. Results of this work showed that listeners were able to use information about strong and weak syllables (when this information was available) to make decisions about word boundaries. Liss et al. (2000) cited this finding as evidence that language processing strategies employed by listeners of dysarthric speakers may not differ from those employed by listeners of nondisordered speakers. However, they suggested that listeners' success in applying normal language processing strategies was directly influenced by acoustic features of the speech signal. Thus, when speech was more degraded or impaired, listeners seemed to be less successful in using typical processing strategies. Recent research suggests that perceptual reorganization may be necessary for listeners to successfully decipher dysarthric speech, especially when vowel space area is reduced (Liu, Tsao, & Kuhl, 2005).

To date, the kinds of linguistic errors that listeners make when faced with dysarthric speech have not been characterized. This linguistic perspective is important because it can provide insight into the extent to which listeners are successful in correctly identifying different types of words that vary in importance from a semantic perspective. For this purpose, content words (nouns and verbs) could be considered to be of primary importance because they carry key information. Modifiers (adjectives and adverbs) could be considered to be of secondary importance because they serve to qualify content words. Functors (articles and prepositions) could be considered to be of less importance because their primary role is structural or syntactic; thus, they carry little meaning in and of themselves. This type of linguistic class analysis would describe the

nature of the linguistic problems that listeners experience when faced with dysarthric speech as well as the potential impact of such problems on the exchange of meaning between speaker and listener. Further, characterization of linguistic errors made by listeners may have implications for interventions that target listener-based perceptual strategies for understanding dysarthric speech. Toward this end, the second goal of the present study was to examine linguistic errors made by listeners of speakers with dysarthria to determine whether particular types of words were more vulnerable to misperception than others.

The following specific research questions were addressed:

1. Do three paradigms for scoring orthographic transcriptions of sentences produced by speakers with dysarthria yield different intelligibility results?
2. Are listeners more accurate in transcribing words from certain linguistic classes than others produced by speakers with dysarthria?
3. Are listeners more accurate in transcribing bound versus free grammatical morphemes produced by speakers with dysarthria?

Method

Participants

Two groups of participants were involved in this research: speakers with dysarthria and listeners with normal communication abilities.

Speakers. Twelve adults with dysarthria secondary to cerebral palsy served as speakers. They were selected to represent a range of intelligibility levels. Table 1 provides demographic information on the speakers. 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 (Hustad & Beukelman, 2002) between 5% and 95%, (d) be between 18 and 60 years old, (e) be able to produce connected speech consisting of at least eight consecutive words (with no breath group or time limitations), and (f) be able to repeat sentences of up to eight words in length following a verbal model.

Speakers were assigned retrospectively to severity groups to simplify statistical analyses. Severity groupings were made based on intelligibility scores obtained for the speech stimuli employed in the study. Three speakers were assigned to each of the four severity groups as follows: speakers with narrative transcription intelligibility scores between 75% and 95% were in the mild group, between 50% and 70% were in the moderate group, between 25% and 45% were in the severe group, and between 5% and 20% were in the profound group.

Listeners. Twelve different individuals listened to speech stimuli for each of the 12 speakers, for a total of 144 listeners (23 men, 121 women). Different listeners were randomly assigned to each speaker so that the same stimulus material could be used for each speaker without the possibility of a learning effect. This type of paradigm is frequently used

TABLE 1. Demographic information for speakers with dysarthria.

Speaker	Age (years)	Sex	Dysarthria diagnosis	Dysarthria severity	Perceptual features of connected speech	Transcription intelligibility score (%)
1	37	M	Spastic	Profound	Harsh vocal quality, imprecise articulation, short phrases	5
2	33	M	Hyperkinetic-spastic	Profound	Irregular articulatory breakdowns, imprecise articulation	15
3	24	F	Hyperkinetic-spastic	Profound	Irregular articulatory breakdowns, imprecise articulation	16
4	58	F	Spastic	Severe	Hypernasality, imprecise articulation	30
5	46	F	Spastic	Severe	Harsh vocal quality, hypernasality, imprecise articulation	37
6	42	F	Spastic	Severe	Hypernasality, imprecise articulation	39
7	21	M	Hyperkinetic-spastic	Moderate	Harsh vocal quality, irregular articulatory breakdowns, imprecise articulation	58
8	33	F	Spastic	Moderate	Breathy voice, short phrases, imprecise articulation	59
9	55	M	Spastic	Moderate	Hypernasality, imprecise articulation, short phrases	65
10	37	M	Spastic	Mild	Harsh vocal quality, imprecise articulation, short phrases	80
11	32	F	Spastic	Mild	Breathy voice, imprecise articulation, short phrases	86
12	53	F	Hyperkinetic-spastic	Mild	Hypernasality, imprecise articulation, short phrases	86

Note. Transcription intelligibility scores are the same as those used for the total word phonemic match scoring paradigm. Dysarthria type, severity, and prominent perceptual features were identified on the basis of medical records and clinical assessment by the author, a certified speech-language pathologist.

in transcription intelligibility studies (see Beukelman, Fager, Ullman, Hansen, & Logemann, 2002; Garcia & Dagenais, 1998; Hustad & Cahill, 2003; Hustad et al., 2003). 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 HL 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 old; and (e) have no identified language, learning, or cognitive disabilities per self-report. All listeners were either currently attending college or graduate school or had completed college or graduate school. Listeners had a mean age of 21.25 years ($SD = 2.43$).

Materials

Speakers with dysarthria produced three narrative passages, each constructed of 10 related sentences. The passages employed in this study have been used in other projects focused on intelligibility of dysarthric speech (see Hustad & Beukelman, 2001, 2002; Hustad et al., 2003). In summary, passages were developed to represent common situations (e.g., sporting event, natural disaster, purchasing a vehicle). Passages followed standard American English conventions for content, form, and use of the language. Each of the narrative passages contained a total of 65 words. The 10 constituent sentences systematically ranged in length from 5 to 8 words. Each narrative contained 49 different words, had a type:token ratio of 0.75, and was at a fifth-grade reading level. Passages and constituent sentences were not systematically controlled for syntactic structure or phonetic composition.

Procedures

Recording speech samples. Speech samples from each of the 12 speakers were recorded on digital audiotape (48-kHz sampling rate, 16-bit quantization) while the speakers produced the target narratives. All recordings took place individually in a quiet environment, either in the speaker's home or in a sound-attenuating room in the laboratory. Speakers wore a unidirectional head-mounted microphone positioned 5 cm from the mouth. To ensure that differences in reading fluency and visual acuity did not affect the production of target sentences, speakers produced the individual sentences from each narrative following the experimenter's model. Orthographic representations of stimulus sentences were also provided on a computer screen positioned in front of the speakers. Speakers were required to produce each sentence verbatim, including all constituent words, and were asked to repeat any sentence that did not include all words per the experimenter's perceptual judgment. Repetitions were required on fewer than 5% of stimulus sentences across all speakers. Speakers were encouraged to speak naturally, as they would in real communication situations.

Preparing speech samples for playback. Recorded samples were transferred onto computer via a digital sound card, maintaining the sampling rate and quantization of the original recordings. For each speaker, recordings of each stimulus sentence were separated into individual sound files. While digitizing the files, a research assistant listened to each sentence and confirmed that each target word from each sentence was present. Stimulus files for each sentence were normalized using Sound Forge 6.0 (Sonic Foundry, 2002)

so that the peak amplitude of each sentence was constant across all files. Root-mean-square (RMS) amplitude was also determined for all sentences produced by each speaker to ensure that stimuli were similar in average intensity. Across speakers and sentences, the average RMS value was 64.89 ($SD = 2.78$) for speakers in the mild group, 66.59 ($SD = 2.29$) for speakers in the moderate group, 65.80 ($SD = 1.86$) for speakers in the severe group, and 64.86 ($SD = 3.49$) for speakers in the profound group.

Randomization and counterbalancing. Each listener transcribed 10 sentences from one narrative passage produced by one speaker. Presentation of the three different narrative passages was counterbalanced so that each of the three passages was represented an equal number of times among the listeners of each speaker.

Listening task. Listeners completed the experiment independently in a soundproof booth. Each listener was seated approximately 2 ft from a high-quality external speaker, with a desktop computer located directly in front of him or 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.

All experimental tasks were presented via computer using a custom programmed setup in Microsoft PowerPoint. Before beginning the experimental tasks, listeners were instructed that they would hear a person with a speech problem who would be producing a short story consisting of 10 sentences. They were told to type exactly what they thought the speaker said, during a pause between each sentence. They were informed that they could hear each sentence only one time and that the person speaking would be difficult to understand. Listeners were instructed to venture their best guess if they were uncertain of what the speaker said. They were informed that they could take as much time as necessary and that the experiment was completely self-paced.

Scoring

Scoring paradigms. Listener-generated orthographic transcriptions of speakers with dysarthria were evaluated using three different scoring paradigms, each of which yielded a “percentage correct” value that was used to compare differences among paradigms. It is important to note that the same orthographic transcriptions were used for the different scoring paradigms.

The first paradigm, total word phonemic match (TPM), was used to characterize the extent to which listener transcriptions were an exact phonemic match to the target utterances produced by the speakers. This paradigm employed standard procedures used in other transcription intelligibility studies (see Garcia & Dagenais, 1998; Hustad et al., 2003). To score TPM, each transcribed word was evaluated to determine whether it was an exact phonemic match with the target word. Misspellings and homonyms were accepted as correct, as long as all *phonemes* in the spoken version of the transcribed word were correct. For example, if a listener wrote the word *they're* in place of the target word *there*, the word would be counted as correct. Each word within the listener-generated orthographic transcripts was counted as

either correct or incorrect. The number of words identified correctly was tallied and divided by the number of words possible. The resultant proportion was used for analyses.

The second scoring paradigm, informational word phonemic match (IPM), was used to characterize the extent to which listeners were able to transcribe only information-bearing words as an exact phonemic match with the target words produced by speakers with dysarthria. Words considered in the measurement of IPM were all content words (defined as all nouns and verbs) and all modifiers (defined as adverbs and adjectives). IPM was scored exactly the same as TPM except that all functor words (prepositions, articles, conjunctions) were excluded from the count of words identified correctly. For analyses, all content words and modifiers identified correctly were tallied and divided by the number of informational words possible.

The third scoring paradigm, informational word semantic match (ISM), was also used to characterize the extent to which listeners were able to transcribe only the information-bearing words produced by speakers. However, ISM focused on determining the extent to which listener transcriptions of information-bearing words reflected the semantic intent of the target words, not necessarily the precise form. Accordingly, the criterion for a word to be considered correct was considerably less stringent than for the TPM and IPM paradigms. Again, words considered in the measurement of ISM were all content words (defined as all nouns and verbs) and all modifiers (defined as adverbs and adjectives), with functor words excluded. To score ISM, each transcribed informational word was evaluated relative to the target word produced by the speakers to determine whether the transcribed word expressed the same general meaning as the target word. In making these judgments, morphological errors were not regarded negatively; thus, errors involving, for example, number or tense did not cause a word to be counted as incorrect. For analyses, all content words and modifiers that correctly represented the semantic intent of the target word were tallied and divided by the number of informational words possible. See Table 2 for an example of how one transcript was scored using each of the three paradigms.

Linguistic classes. Listener-generated orthographic transcriptions of speakers with dysarthria were also coded for the types of words that listeners transcribed correctly using the TPM paradigm. Words identified correctly were separated into three mutually exclusive linguistic categories. The first category was content words, which included all nouns, pronouns, and verbs. The second category was modifiers, which included all adjectives and adverbs. The third category was functors, which included all articles, prepositions, and conjunctions. Words within each category that were identified correctly for each speaker were tallied and then divided by the number of words possible for each category. This proportion was used to compare differences in transcription accuracy among linguistic categories.

Morphemes. Orthographic transcriptions were also coded for grammatical morphemes transcribed correctly, again using the TPM paradigm. Free morphemes (i.e., morphemes that could stand alone as a word) and bound inflectional grammatical morphemes (i.e., plural, possessive, regular tense

TABLE 2. Sample listener transcript scored using each of the three scoring paradigms.

Target sentence	Listener transcript	Scoring with TPM	Scoring with IPM	Scoring with ISM
Jason needed to buy a car.	Jason is standing by a corner.	Jason by a	Jason by	Jason
He wanted a new car.	He wanted a new car.	He wanted a new car	He wanted new car	He wanted new car
He considered two different models.	He considered two different models.	He considered two different models	He considered two different models	He considered two different models
Four-wheel drive was a desired feature.	Four-wheel drive and different features.	Four-wheel drive	Four-wheel drive	Four-wheel drive features
Jason liked the large pickup trucks.	Jason looked at the large prices.	Jason the large	Jason large	Jason large
Sport-utility vehicles were his favorite.	But utility vehicles were his favorite.	Utility vehicles were his favorite	Utility vehicles were his favorite	Utility vehicles were his favorite
He did not have much money to spend.	He did not have much money to spend.	He did not have much money to spend	He did not have much money spend	He did not have much money spend
He bargained with a salesman for 2 hr.	He bargained with the salesman for 2 hr.	He bargained with salesman for 2 hr	He bargained salesman 2 hr	He bargained salesman 2 hr
The final price was within his budget.	The sale price was in his budget.	The price was his budget	Price was his budget	Sale price was his budget
A used Jeep was what he purchased.	A new Jeep was what he purchased.	A jeep was what he purchased	Jeep was what he purchased	Jeep was what he purchased

Note. TPM = total word phonemic match; IPM = informational word phonemic match; ISM = informational word semantic match.

inflections) that were transcribed correctly for each speaker were tallied and divided by the number of free morphemes possible and bound morphemes possible, respectively. Again, this proportion was used to compare difference in transcription accuracy between the two types of morphemes.

Reliability

Intra- and interscorer reliability was determined for each of the dependent measures: TPM, IPM, and ISM. In addition, inter- and intracoder reliability was determined for the classification of correctly transcribed words as functors, modifiers, or content words, and as bound and free grammatical morphemes. Data for 3 randomly selected listeners from each of the 12 listener groups (25% of the sample) were rescored for inter- and intrajudge reliability. Point-by-point reliability was obtained by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100.

Interscorer reliability was 99% for TPM, 97.20% for IPM, and 95.75% for ISM. Intrascorer reliability was 99% for TPM, 99.05% for IPM, and 97.29% for ISM. Intercoder linguistic classification reliability was 98.75% across all linguistic classes and the two types of grammatical morphemes. Intrarater linguistic classification reliability was 98.52%. These findings document strong intra- and interjudge reliability in the scoring and classifying of dependent measures.

Experimental Design and Statistical Procedures

Three separate sets of analyses were employed for this study. Split-plot designs and fully factorial analyses of variance (ANOVAs; Kirk, 1995) were used for each analysis, with the between-subjects measure in each analysis being

severity (mild, moderate, severe, profound). Three speakers composed each severity group, and 12 listeners heard each speaker, for a total of 36 different listeners per severity group.

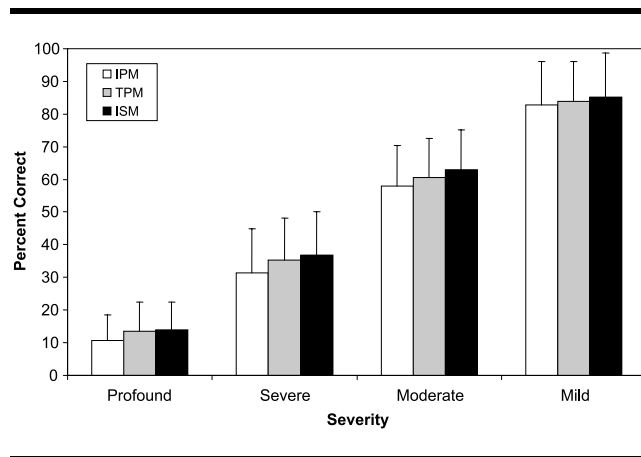
The first analysis examined differences among scoring paradigms used to evaluate listener transcriptions. The within-subjects repeated variable was measure, and its three categories were TPM, IPM, and ISM. The second analysis evaluated differences among linguistic classes of words that were transcribed correctly. The within-subjects repeated variable was linguistic class, and its three categories were functors, content words, and modifiers. The final analysis evaluated the differences between the number of bound and free grammatical morphemes that were transcribed correctly. This analysis was conducted separately from the linguistic class analysis because grammatical morphemes were not mutually exclusive from the other linguistic classes. The within-subjects repeated variable was morpheme, and its two categories were bound and free.

Results

Differences Among Scoring Paradigms

Descriptive statistics, illustrated in Figure 1, suggested lower average intelligibility scores for IPM, and similar scores for TPM and ISM for each severity group. ANOVA results, using the Greenhouse-Geisser adjustment, showed that the main effect of scoring procedure was significant, $F(1.65, 231.37) = 101.24, p < .0001, \eta^2 = .420$. The main effect of severity was also significant, $F(3, 140) = 248.57, p < .0001, \eta^2 = .842$. Finally, the interaction between severity and scoring procedure was significant, $F(4.95, 231.37) = 3.78, p = .003, \eta^2 = .08$.

FIGURE 1. Intelligibility by scoring method and severity (TPM = total word phonemic match; IPM = informational word phonemic match; ISM = informational word semantic match). Error bars represent +1 SD of listener performance.



A series of 15 paired-samples *t* tests, with alpha partitioned using the Bonferroni procedure, were used to examine follow-up contrasts of interest (Howell, 2002, 2004; Marascuilo & Levin, 1983; Marascuilo & Serlin, 1988). Findings, presented in Table 3, showed that when data from all speakers were pooled, ISM resulted in higher intelligibility scores than IPM, TPM resulted in higher intelligibility scores than IPM, and ISM resulted in higher intelligibility scores than TPM. Results within the different severity groups showed a somewhat different pattern of results, with one exception: ISM resulted in higher intelligibility scores than IPM for all severity groups. For the moderate, severe, and profound groups, TPM resulted in higher intelligibility scores than IPM. For the moderate group, ISM resulted in higher intelligibility scores than TPM.

Differences Among Linguistic Classes

Descriptive statistics, illustrated in Figure 2, suggested higher average intelligibility scores for functor words and similar scores for content words and modifiers for each severity group. ANOVA results showed that the main effect of linguistic class was significant, $F(1.57, 220.47) = 96.40$, $p < .0001$, $\eta^2 = .408$. The main effect of severity was also significant, $F(3, 140) = 218.13$, $p < .0001$, $\eta^2 = .824$. Finally, the interaction between severity and linguistic class was significant, $F(4.72, 220.47) = 3.92$, $p = .002$, $\eta^2 = .078$.

Following procedures described above, 15 paired-samples *t* tests were performed to examine follow-up contrasts of interest. Findings, presented in Table 4, showed that when data from all speakers were pooled, functor words were significantly more intelligible than modifiers and content words. These same results occurred within the moderate, severe, and profound groups as well. For the mild group, the only significant finding was that functor words were more intelligible than content words.

Differences Among Types Of Morphemes

Descriptive statistics, illustrated in Figure 3, suggested lower average intelligibility scores for bound morphemes

TABLE 3. Follow-up contrasts comparing results of different scoring procedures within severity groups.

Contrast	Mean difference (% correct)	df	SE	<i>t</i>	Observed <i>p</i> value
All speakers					
ISM vs. TPM	1.39	143	0.35	3.97	<.0001*
IPM vs. TPM	-2.67	143	0.28	-9.65	<.0001*
IPM vs. ISM	-4.07	143	0.26	-15.62	<.0001*
Mild dysarthria					
ISM vs. TPM	1.31	35	0.74	1.74	.0850
IPM vs. TPM	-1.17	35	0.63	-1.85	.0730
IPM vs. ISM	-2.47	35	0.49	-5.01	<.0001*
Moderate dysarthria					
ISM vs. TPM	2.42	35	0.70	3.43	.0020*
IPM vs. TPM	-2.57	35	0.49	-5.26	<.0001*
IPM vs. ISM	-4.99	35	0.57	-8.71	<.0001*
Severe dysarthria					
ISM vs. TPM	1.41	35	0.69	2.02	.0510
IPM vs. TPM	-3.96	35	0.52	-7.69	<.0001*
IPM vs. ISM	-5.36	35	0.46	-11.59	<.0001*
Profound dysarthria					
ISM vs. TPM	0.45	35	0.66	.67	.5090
IPM vs. TPM	-2.99	35	0.49	-6.15	<.0001*
IPM vs. ISM	-3.44	35	0.41	-8.44	<.0001*

* $p < .003$ (.05/15 contrasts).

than for free morphemes within each severity group. ANOVA results showed that the main effect of morpheme type was significant, $F(1, 140) = 400.78$, $p < .0001$, $\eta^2 = .741$, with free morphemes being more intelligible than bound morphemes across speaker severity groups (mean difference = 18.73%). The main effect of severity was also significant, $F(3, 140) = 253.23$, $p < .0001$, $\eta^2 = .844$. Finally, the interaction between severity and morpheme type was significant, $F(3, 140) = 8.69$, $p < .0001$, $\eta^2 = .157$.

Following the same procedures as described previously, four paired-samples *t* tests were performed to examine follow-up contrasts of interest. Results, shown in Table 5, revealed that free morphemes were significantly more intelligible than bound morphemes within each of the four severity groups.

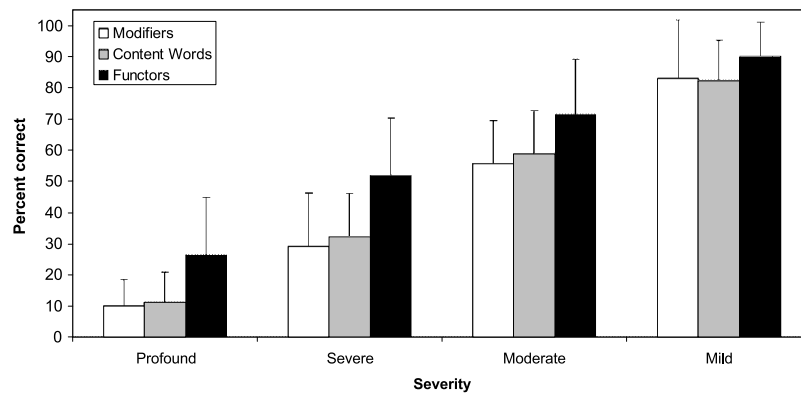
Discussion

This study examined whether different paradigms for scoring orthographic transcriptions of speakers with dysarthria influenced intelligibility results. Also of interest was whether there were differences in transcription accuracy for words from different linguistic classes (i.e., content words, modifiers, and functors) and for different morphological categories (i.e., bound vs. free grammatical morphemes). This discussion focuses on interpretation of the findings and clinical implications for each of these questions.

Scoring Paradigms

Results of this study indicated that different paradigms for scoring listener-generated orthographic transcriptions of

FIGURE 2. Intelligibility by linguistic class and severity (modifiers = adjectives and adverbs; functors = articles, prepositions, conjunctions; content words = nouns, pronouns, and verbs). Error bars represent +1 SD of listener performance.



speakers with dysarthria did, in fact, yield statistically different results. Not surprisingly, when only informational words were scored using a more liberal semantic paradigm (ISM), intelligibility scores tended to be higher than when the analogous exact phonemic match procedure was employed to evaluate those same informational words (IPM). This difference was significant within each severity group. Although the magnitude was relatively small (2.47%–5.36%), this finding indicates that listeners were able to interpret the meaning of words with greater accuracy than

they were able to decode the precise form of the same words. One conclusion is that listeners seek and obtain content more readily than they seek and obtain exact form when presented with dysarthric speech. This would seem to have important communicative consequences for facilitating the exchange of meaning between speaker and listener.

Another finding was that intelligibility scores were higher for the exact phonemic match paradigm when all words were scored (TPM) than when only informational words were scored (IPM). This group finding was true for speakers within all severity groups except for the mild group. One explanation can be found in the analysis of the linguistic classes of words that listeners transcribed correctly. Results clearly showed that listeners consistently transcribed non-information-bearing functor words with greater accuracy than any other word class (except for listeners of speakers with mild dysarthria). Thus, it seems logical that the higher intelligibility scores associated with the TPM paradigm could be attributed to the disproportionate number

TABLE 4. Follow-up contrasts comparing errors among linguistic classes within severity groups.

Contrast	Mean difference (% correct)	df	SE	t	Observed p value
All speakers					
Modifiers vs. content	-1.71	143	0.96	-1.77	.0790
Functors vs. content	13.85	143	1.20	11.49	<.0001*
Modifiers vs. functors	-15.56	143	1.56	-9.97	<.0001*
Mild dysarthria					
Modifiers vs. content	.77	35	1.81	.43	.6710
Functors vs. content	7.75	35	1.73	4.49	<.0001*
Modifiers vs. functors	-6.97	35	2.45	-2.85	.0070
Moderate dysarthria					
Modifiers vs. content	-3.22	35	2.25	-1.43	.1610
Functors vs. content	12.68	35	2.40	5.28	<.0001*
Modifiers vs. functors	-15.89	35	3.61	-4.39	<.0001*
Severe dysarthria					
Modifiers vs. content	-3.14	35	1.81	-1.74	.0910
Functors vs. content	19.69	35	2.65	7.42	<.0001*
Modifiers vs. functors	-22.84	35	2.84	-8.05	<.0001*
Profound dysarthria					
Modifiers vs. content	-1.23	35	1.79	-.69	.4960
Functors vs. content	15.28	35	2.41	6.35	<.0001*
Modifiers vs. functors	-16.52	35	2.99	-5.51	<.0001*

*p < .003 (.05/15 contrasts).

FIGURE 3. Intelligibility by grammatical morpheme type and severity. Error bars represent +1 SD of listener performance.

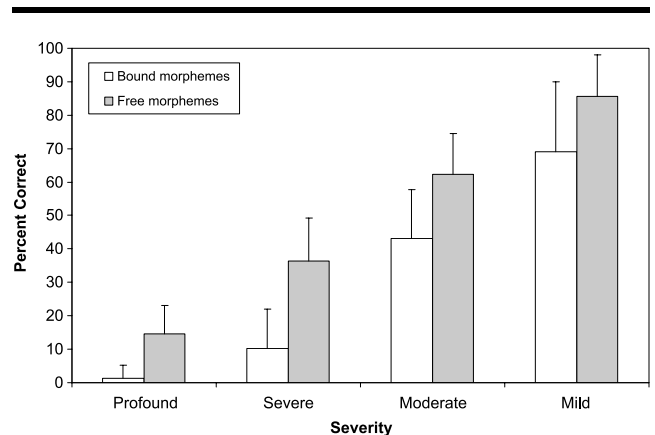


TABLE 5. Follow-up contrasts comparing free and bound grammatical morphemes identified correctly within speaker severity group.

Contrast	Mean difference (% correct)	df	SE	t	Observed p value
Free vs. bound (mild dysarthria)	16.54	35	2.29	7.21	<.0001*
Free vs. bound (moderate dysarthria)	19.14	35	1.94	9.86	<.0001*
Free vs. bound (severe dysarthria)	26.14	35	1.65	15.82	<.0001*
Free vs. bound (profound dysarthria)	13.12	35	1.49	8.76	<.0001*

* $p < .0125$ (.05/4 contrasts).

of functor words that were transcribed correctly and were weighted equally relative to the other, more important, words. In the case of speakers with moderate, severe, and profound dysarthria, this equal weighting may actually overrepresent listeners' ability to decode correctly the most important words produced by speakers. A caveat, however, is that the magnitude of the difference between scores obtained with IPM versus TPM was relatively small, with differences ranging from 2.57% to 3.96%. One important reason for the small magnitude of this difference is that functor words occurred less frequently than other types of words within the target utterances produced by speakers (19% of words were functors; 81% of words were content words or modifiers).

Interestingly, semantic scoring of information-bearing words (ISM) and scoring of all words using the exact phonemic match paradigm (TPM) tended to yield similar results for speakers in the mild, severe, and profound groups. For the speakers in the moderate group, there was a small but significant advantage to the semantic scoring (ISM). In general, this finding suggests that although the disproportionate number of functor words identified correctly may inflate the TPM score, this inflation actually provides a reasonable estimation of the percentage of information-bearing words that listeners *interpreted* correctly but did not necessarily decode perfectly. From a clinical perspective, findings suggest that the use of a rigid phonemic match procedure or a more flexible semantic match procedure may not be of important consequence as long as a consistent scoring paradigm is used for repeated measures on the same patient.

Linguistic Classes and Morphemes

Results of the present study showed very clear differences in the accuracy with which listeners were able to transcribe words from different linguistic classes. Specifically, listeners were able to transcribe functor words (conjunctions, articles, prepositions) with a significantly higher degree of accuracy than content words (nouns and verbs) and modifiers (adverbs and adjectives). The magnitude of this difference was remarkable, ranging between approximately 8% and 23%. One reason that functor words may have been easier to transcribe than other types of words relates to their

phonetic and syllabic structure. Functor words tend to be monosyllabic CV or VC words (e.g., *in, on, an, the, of*); thus, they may be easier for speakers to produce accurately and consequently easier for listeners to perceive accurately. Another explanation relates to the predictability of these words with regard to their syntactic role in sentences. That is, functor words tend to be highly predictable within the syntactic structure of sentences. As such, top-down linguistic knowledge may have played an important role in helping listeners identify these words. Conversely, content words and modifiers tend to be more complex in their phonetic structure and their syllabic structure, making them more difficult to produce. In addition, these words are also less predictable and are considered open class in nature, which may make it more difficult for listeners to use top-down knowledge to aid in identifying these words.

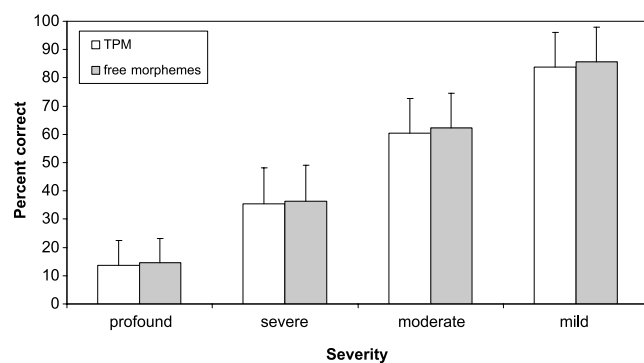
The findings of the present study provided no information regarding speaker production performance, beyond what can be inferred from listener perception performance. Turner and Tjaden (2000) examined production differences between content and functor words produced by speakers with dysarthria secondary to amyotrophic lateral sclerosis. Their findings suggested that for speakers with dysarthria, vowel durations tended to be shorter and the first and second formant frequency values tended to be more centralized for functor words relative to content words. Their findings were generally consistent with research regarding acoustic differences between content and functor words for normal speakers (Pichney, Durlach, & Braida, 1986). Grosjean and Gee (1987) suggested that these acoustic differences between content and functor words may serve to add emphasis to the more important information-bearing words within the speech signal and help listeners to focus their processing on the most important words. However, because Turner and Tjaden (2000) did not collect transcription intelligibility data from listeners, the extent to which their findings were consistent with listener perception is not known. Results from the present study may suggest that listeners of speakers with dysarthria are not able to use these acoustic cues to aid in their processing of dysarthric speech. Another explanation is that knowledge intrinsic to the listener (i.e., top-down knowledge of the language, world knowledge, predictability of words) may play a more important role in processing dysarthric speech than production adjustments made by speakers. Finally, it is possible that speakers with dysarthria secondary to cerebral palsy do not mark differences acoustically between content and functor words in the same way as speakers with amyotrophic lateral sclerosis. Further research is necessary to determine the extent to which production features of speakers with cerebral palsy map onto listener perception at multiple linguistic levels including the sound segment, the word, and the entire sentence.

With regard to grammatical morphemes, results of the present study showed that listeners were able to transcribe free morphemes much more readily than bound grammatical morphemes. This is not especially surprising given that bound morphemes often take the form of a single phoneme or syllable added to the end of a word. Indeed, research has shown that speakers with cerebral palsy tend to make more

production errors on word-final consonants than word-initial consonants (Platt, Andrews, & Howie, 1980). In the present study, speakers were required to produce an approximation of each target word, and this was verified by two different sources (see Method section). However, many speakers in the study had significantly compromised intelligibility, which made it difficult to verify that they produced specific *phonemes* or phoneme approximations (i.e., bound morphemes). In many cases, word, syllable, or phoneme approximations were so distorted that their accuracy could not be confirmed by the experimenters. An alternative, or perhaps complementary, explanation relates to the relative importance of grammatical morphemes to the meaning of individual words. Because bound morphemes do not necessarily make a critical contribution to the content of a word or sentence and because they are likely to be more difficult for speakers to produce, it is possible that listeners did not attend to them or allocated their processing resources to the more important information-bearing free morphemes. Additional qualitative and quantitative study is necessary to characterize the quality of speakers' productions of bound morphemes and to relate production characteristics with perception data such as those presented in this study.

Although listeners clearly had difficulty with bound morphemes, this did not seem to have important consequences for intelligibility scores. A descriptive comparison is shown in Figure 4. Here, scores for the exact phonemic match procedure (TPM) in which all morphemes were required to be correctly transcribed were compared with a procedure such as that employed by Liss et al. (2002) that essentially counted free morphemes (i.e., errors on bound morphemes did not penalize the intelligibility score). Although not analyzed statistically, data suggest that the differences between these procedures would not likely be clinically meaningful (less than 2% difference within each severity group). An important reason that this difference was very small is that the vast majority of the morphemes within the target stimuli were free morphemes (88% free, 12% bound); thus, there was opportunity on only a small proportion of the words within the sample for an error on a bound morpheme to render a word incorrect.

FIGURE 4. Comparison of intelligibility for TPM versus free morphemes by severity.



Conclusions

Results of the present study showed that different paradigms for scoring listener-generated orthographic transcriptions of speakers with dysarthria—for example, exact phonemic match for all words, exact phonemic match for information-bearing words, and semantic match for information-bearing words—result in different intelligibility scores. However, the magnitude of the differences among paradigms, although statistically significant, tended to be small and would probably not be considered clinically significant. It is important to note that listeners in this study were naive. In clinical practice, individuals who listen to and transcribe speech samples may vary in their level of experience with dysarthric speech. As a result, the impact of scoring procedures on intelligibility measures may be somewhat different than those observed in this study. Nonetheless, when making clinical intelligibility measures, it is important that clinicians use consistent scoring procedures, particularly with the same patient, because small differences among paradigms could have an important influence on interpretation of, for example, progress associated with treatment.

The present study also demonstrated that there were important differences in the types of words that listeners were able to transcribe correctly. Words that had simpler syllabic and phonemic structure and were more predictable from a syntactic perspective (i.e., functor words) were easier for listeners to transcribe correctly than other, more important, information-bearing words such as content words and modifiers. Liss et al. (2002) suggested that listeners of speakers with dysarthria may apply similar perceptual strategies for processing both normal and disordered speech. However, the effectiveness of these strategies is affected by the nature and severity of the speech impairment. The results of the present study suggest that it might be useful to teach listeners alternative processing strategies to enhance their ability to decode information-bearing words. To help listeners focus on important content-related words, it might be of value to instruct listeners that functor words may be easiest for them to understand and that they should explicitly direct their attention to other types of words to help facilitate successful communication. It may also be helpful for listeners to engage in explicit comprehension-monitoring activities such as shadowing (Hustad & Shapley, 2003) and requesting clarification as soon as they have difficulty interpreting information-bearing words. Further research is necessary to identify and evaluate the feasibility and usefulness of these types of listener-directed strategies.

Acknowledgments

This research was funded by National Institute on Deafness and Other Communication Disorders Grant R03 DC005536. The author thanks Caitlin Dardis and Lisa Igl for assistance with coding orthographic transcriptions.

References

- Ansel, B. M., & Kent, R. D. (1992). Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy. *Journal of Speech and Hearing Research, 35*, 296–308.

- Beukelman, D. R., Fager, S., Ullman, C., Hanson, E., & Logemann, J.** (2002). The impact of speech supplementation and clear speech on the intelligibility and speaking rate of people with traumatic brain injury. *Journal of Medical Speech Language Pathology, 10*, 237–242.
- 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.
- Grosjean, F., & Gee, J. P.** (1987). Prosodic structure and spoken word recognition. In U. Frauenfelder & L. K. Tyler (Eds.), *Spoken word recognition* (pp. 135–155). Amsterdam, The Netherlands: Elsevier Science.
- Howell, D. C.** (2002). *Statistical methods for psychology* (5th ed.). Belmont, CA: Duxbury Press.
- Howell, D. C.** (2004). *Fundamental statistics for the behavioral sciences* (5th ed.). Belmont, CA: Duxbury Press.
- 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.** (2002). Listener comprehension of severely dysarthric speech: Effects of linguistic cues and stimulus cohesion. *Journal of Speech, Language, and Hearing Research, 45*, 545–558.
- Hustad, K. C., & Cahill, M. A.** (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology, 12*, 1–11.
- Hustad, K. C., Jones, T., & Dailey, S.** (2003). Implementing speech supplementation strategies: Effects on intelligibility and speech rate of individuals with chronic severe dysarthria. *Journal of Speech, Language, and Hearing Research, 46*, 462–474.
- Hustad, K. C., & Shapley, K. L.** (2003). AAC and the development of natural speech. In J. Light, J. Reichle, & D. Beukelman (Eds.), *Augmentative and alternative communication for individuals with developmental disabilities* (pp. 41–62). Baltimore: Brookes.
- Kirk, R.** (1995). *Experimental design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.
- Liss, J. M., Spitzer, S., Caviness, J. N., & Adler, C.** (2002). Effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *Journal of the Acoustical Society of America, 112*, 3022–3030.
- Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., & Edwards, B.** (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *Journal of the Acoustical Society of America, 104*, 2457–2466.
- Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., & Edwards, B. W.** (2000). Lexical boundary error analysis in hypokinetic and ataxic dysarthria. *Journal of the Acoustical Society of America, 107*, 3415–3424.
- Liu, H., Tsao, F., & Kuhl, P.** (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *Journal of the Acoustical Society of America, 117*, 3879–3889.
- Marascuilo, L., & Levin, J.** (1983). *Multivariate statistics in the social sciences: A researcher's guide*. Monterey, CA: Brooks/Cole.
- Marascuilo, L., & Serlin, R.** (1988). *Statistical methods for the social and behavioral sciences*. New York: Freeman.
- Monsen, R. B.** (1978). Toward measuring how well hearing-impaired children speak. *Journal of Speech and Hearing Research, 21*, 197–219.
- Monsen, R. B.** (1983). The oral speech intelligibility of hearing-impaired talkers. *Journal of Speech and Hearing Disorders, 43*, 286–296.
- Pichney, M. A., Durlach, N. I., & Braid, L. D.** (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29*, 434–446.
- Platt, L. J., Andrews, G., & Howie, P.** (1980). Dysarthria of adult cerebral palsy: II. Phonemic analysis of articulation errors. *Journal of Speech and Hearing Research, 23*, 41–55.
- Sonic Foundry.** (2002). Sound Forge 6.0 [Computer software]. Madison, WI: Author.
- Tikofsky, R. S., & Tikofsky, R. P.** (1964). Intelligibility as a measure of dysarthric speech. *Journal of Speech and Hearing Research, 7*, 325–333.
- Turner, G. S., & Tjaden, K.** (2000). Acoustic differences between content and function words in amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research, 43*, 769–781.
- Yorkston, K., & Beukelman, D.** (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders, 11*, 499–512.
- 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., Strand, E. A., & Bell, K. R.** (1999). *Management of motor speech disorders in children and adults* (2nd ed.). Austin, TX: Pro-Ed.

Received May 27, 2005

Revision received October 22, 2005

Accepted March 5, 2006

DOI: 10.1044/1058-0360(2006/025)

Contact author: Katherine C. Hustad, 475 Waisman Center, University of Wisconsin—Madison, 1500 Highland Avenue, Madison, WI 53705. E-mail: kchustad@wisc.edu

Copyright of *American Journal of Speech-Language Pathology* is the property of American Speech-Language-Hearing Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.