Published in final edited form as:
J Speech Lang Hear Res. 2008 June ; 51(3): 562–573. doi:10.1044/1092-4388(2008/040).

The relationship between listener comprehension and intelligibility scores for speakers with dysarthria

Katherine C. Hustad
Department of Communicative Disorders & Waisman Center University of Wisconsin – Madison Madison, WI

Abstract

Purpose—This study examined the relationship between listener comprehension and intelligibility scores for speakers with mild, moderate, severe, and profound dysarthria. Relationships were examined across all speakers and their listeners when severity effects were statistically controlled, within severity groups, and within individual speakers with dysarthria.

Method—Speech samples were collected from 12 speakers with dysarthria secondary to cerebral palsy. For each speaker, 12 different listeners completed two tasks (for a total of 144 listeners), one task involved making orthographic transcriptions and one task involved answering comprehension questions. Transcriptions were scored for the number of words transcribed correctly; comprehension questions were scored on a 3-point scale according to their accuracy.

Results—Across all speakers, the correlation between comprehension and intelligibility scores was non-significant when the effects of severity were factored out and residual scores were examined. Within severity groups, the relationship was significant only for the mild group. Within individual speaker groups, the relationship was non-significant for all but two speakers with dysarthria.

Conclusions—Findings suggest that transcription intelligibility scores do not accurately reflect listener comprehension scores. Measures of both intelligibility and listener comprehension may provide a more complete description of the information-bearing capability of dysarthric speech than either measure alone.

Keywords

speech intelligibility; speech perception; dysarthria; cerebral palsy; comprehensibility

The characterization of dysarthric speech is a topic of both clinical and theoretical importance. Although the dysarthrias are a heterogeneous group of speech disorders, one common characteristic is reduced intelligibility (Duffy, 2005; Yorkston, Beukelman, Strand, & Bell, 1999). Reduced intelligibility can have a critical impact on communication abilities, and may limit vocational, educational, and social participation. As a result, quality of life may be greatly diminished.

Intelligibility refers to how well a speaker's acoustic signal can be accurately recovered by a listener. Although this definition seems simple, there are many speaker-related and listener-related variables that can impact how well a speech signal is deciphered. For example, research has shown that message predictability (Garcia & Cannito, 1996), message length

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

(Yorkston & Beukelman, 1981), contextual cues (Hunter, Pring & Martin, 1991), visual-facial information (Hustad & Cahill, 2003), and listener experience (Tjaden & Liss, 1995) each have the potential to affect intelligibility in significant ways. Thus, intelligibility is more complex than its definition may suggest.

There are several different ways that intelligibility can be measured. One method is orthographic transcription of standard speech samples by naïve listeners (see Giolas & Epstein, 1963; Tikofsky & Tikofsky, 1964; Yorkston & Beukelman, 1981; Garcia & Cannito, 1996). In this paradigm, listeners hear a speech sample (usually sentence-length), and then write down what they thought the speaker said. Constituent transcribed words are scored as either correct or incorrect based on whether they match the intended words of the speaker. Percent intelligibility scores are calculated by taking the number of words identified correctly divided by the number of words possible, multiplied by 100. This method is widely used in clinical applications and tools such as the Sentence Intelligibility Test (Yorkston, Beukelman, & Tice, 1996) are available for clinical use. Transcription intelligibility scores provide important information about the integrity of the acoustic signal relative to “normal” (Hustad & Beukelman, 2002), and are often used to describe severity of the dysarthria (Weismer & Martin, 1992). Transcription intelligibility scores can also be used as a basis of comparison to document progress in treatment (Yorkston, et al., 1999). There are, however, several limitations to the information that can be obtained from transcription intelligibility scores. For example, the underlying basis for the intelligibility deficit cannot be determined from an intelligibility score (Kent, Weismer, Kent, & Rosenbek, 1989; Weismer & Martin, 1992). In addition, when individual listener-transcribed words are scored binomially (correct / incorrect) and each word is weighted equally, it is difficult to infer the extent to which listeners interpreted the meaning of the message. In large part, this is because the kinds of words (i.e. content-bearing vs. non-content bearing) transcribed correctly cannot be determined from the intelligibility score alone (Hustad, 2006).

One complementary measure to transcription intelligibility is assessment of listener comprehension, sometimes called comprehensibility in the broader psychology and communication arts literatures. For clarity, it is important to note that in the dysarthria literature, Yorkston, Strand, and Kennedy (1996) have used the term comprehensibility to refer to “contextual intelligibility”, or intelligibility when contextual information is present in different forms, such as semantic cues, syntactic cues, orthographic cues, and gestures. In their measurement of comprehensibility, Yorkston and colleagues employ orthographic transcription and percent correct scores. Thus, comprehensibility as defined by Yorkston and colleagues is a type of intelligibility, with the addition of contextual information.

In contrast, measures of listener comprehension evaluate listeners’ ability to interpret the meaning of messages produced by speakers with dysarthria without regard for accuracy of phonetic and lexical parsing (Hustad & Beukelman, 2002). Listener comprehension can be evaluated by examining listeners’ ability to answer questions about the content of a message or narrative (Hustad & Beukelman, 2002), or by examining listeners’ ability to summarize the content of a narrative passage (Higginbotham, Drazek, Kowarsky, Scally, & Segal, 1994) produced by a speaker.

Theories of discourse psychology provide a basis for conceptualizing different levels of processing that map onto the constructs of intelligibility and listener comprehension of dysarthric speech. Several competing theories exist in which underlying cognitive processes, mechanisms, and architectures differ. However, it is generally accepted that there are multiple levels of representation involved in language processing (Altmann, 2001; Foltz, 2003; Graesser, Millis, & Zwaan, 1997; Singer, 2000; Zwaan & Singer, 2003). van Dijk and Kintsch (1983) proposed that three levels of discourse representation are involved in the

comprehension process, including the surface code, the textbase or proposition, and the situational model. The first level of representation, surface code, refers to "precise word strings" (Singer, 2000; pg. 370), or the exact syntax and morphology of the original message. The intermediate level of representation, textbase or propositional content, refers to the meanings or propositions that are extracted from the surface code, essentially the semantics of the message (Graesser et al., 1997). The highest level of representation, the situation model, reflects the integration of propositions with the world knowledge and the goals of the receiver (Butcher & Kintsch, 2003; Kintsch, 1992; Singer, 2000). In the dysarthria literature, the majority of studies have focused on intelligibility, which can be considered a form of surface code because the focus of measurement is phonetic and lexical identification accuracy. There have been few studies in which propositional content or higher level situation models (comprehension) of dysarthric speech has been examined.

Results of comprehension studies have been equivocal, likely due to methodological differences between studies. For example, Beukelman and Yorkston (1979) examined the relationship between "information transfer" and intelligibility for 9 speakers with dysarthria of varying severity. In this study listeners completed an intelligibility task in which they transcribed a paragraph (in 5-10 word segments) produced by a speaker with dysarthria. Listeners also completed an information transfer task in which they listened to a different paragraph produced by a speaker with dysarthria, and then answered 10 comprehension questions about the content of the paragraph. Percent correct scores were obtained for each task and for each speaker. Results of the Beukelman and Yorkston study showed a strong and significant relationship ($r = .95$) between intelligibility scores and information transfer (comprehension) scores across all speakers. However, Weismer and Martin (1992) identified a key problem related to the confounding effects of severity in Beukelman and Yorkston's study. That is, the correlation between intelligibility and comprehension scores across speakers of varying severity would be strong simply because both measures are correlated with severity. Thus, severity acted as a "third" variable, masking the true relationship that may or may not exist between intelligibility and comprehension scores. Because of this "third variable" effect, the only conclusion that can be drawn from the Beukelman and Yorkston study is that both intelligibility and listener comprehension increase as severity decreases. Weismer and Martin (1992) suggested that proper examination of the relationship between severity-related variables would require that severity be controlled or blocked so that the relationship could be examined within severity groups, rather than across severity groups.

Hustad and Beukelman (2002) examined the relationship between listener comprehension and intelligibility for four speakers with severe dysarthria. They also examined the impact of supplemental contextual cues on the relationship between intelligibility and comprehension. Listeners completed two tasks, one in which they transcribed speech samples produced by a speaker with dysarthria and one in which they answered comprehension questions about a different set of speech samples produced by the same speaker with dysarthria. Results showed a weak and non-significant relationship between intelligibility and comprehension when no cues were provided. This finding provides evidence for the notion that the two measures tap into different phenomena and that listener performance on one measure does not necessarily reflect performance on the other. Thus, listener comprehension measures and speech intelligibility measures appear to provide different, yet complementary, information about the dysarthric speech signal. Because similar studies have not been conducted for groups of speakers from different severity groups, generalization of conclusions regarding the relationship between intelligibility and comprehension is difficult.

The purpose of the present study was to evaluate the relationship between comprehension and intelligibility for listeners of speakers with dysarthria from four different severity

groups. The study was designed to be conceptually similar to that of Beukelman and Yorkston (1979), however, using both experimental and statistical procedures, the effects of severity were controlled (statistically partialled out) and systematically examined. Specifically, this study addressed the following questions: 1.) Across all speakers, when severity effects are partialled out, what is the relationship between intelligibility and comprehension? 2.) Within each speaker severity group, what is the relationship between intelligibility and comprehension? Is this relationship different among the severity groups? 3.) Within individual speakers, what is the relationship between comprehension and intelligibility? What is the extent of individual differences among speakers?

Method

The present study was part of a larger project examining measurement of intelligibility. The first paper (Hustad, 2006) examined the impact of different scoring methods on intelligibility findings. It also examined linguistic class errors made by everyday listeners when orthographically transcribing dysarthric speech. The present study used the same transcription intelligibility data obtained from the same listeners and speakers as reported in the Hustad (2006) paper. However, data were subjected to different analyses, as described below, in the present paper. In addition, data from a separate comprehension task, collected in parallel with intelligibility data, are reported in this paper.

Participants

Speakers with Dysarthria—Twelve speakers with dysarthria secondary to cerebral palsy contributed speech samples for this study. Speakers were selected to represent a range of severity levels. Three speakers were assigned to each of the four severity groups based exclusively on scores from the SIT (Yorkston et al., 1996). For the purposes of this project, severity groups were operationally defined as follows: scores between 76% and 95% were in the mild group; between 46% and 75% were in the moderate group; between 25% and 45% were in the severe group; and between 5% and 24% were in the profound group. Demographic information for the speakers including age, gender, dysarthria diagnosis, severity, prominent perceptual features, intelligibility score, and speech rate is provided in Table 1. Details regarding inclusion criteria can be found in Hustad (2006).

Listeners—Twelve different individuals with normal hearing listened to speech stimuli for each of the 12 speakers with dysarthria, for a total of 144 listeners. Details regarding inclusion criteria for listeners can be found in Hustad (2006). Listeners had a mean age of 21.25 years (SD = 2.43). Twenty-three men and 121 women participated in this study. Sex was not a variable of interest; therefore no attempt was made to balance the number of male and female participants.

Materials

Speech stimuli—Speakers with dysarthria produced 3 narrative passages, each consisting of 10 related sentences. The passages employed in this study have been used in several other projects focused on intelligibility of dysarthric speech (see Hustad & Beukelman, 2001;2002;Hustad, Jones, & Dailey, 2003;Hustad, Auker, Natale, & Carlson, 2003). The interested reader is referred to Hustad and Beukelman (2002) for specific details regarding characteristics of the passages. 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.

Comprehension questions—For each narrative passage, ten comprehension questions were developed. Five questions were designed to be inferential in nature and five were designed to be factual in nature. Inferential questions targeted information that was not overtly specified within the narratives, but could be inferred from the content of the narrative. Factual questions targeted information that was directly stated within the narrative. Because each of the narratives was different, it was not possible to use all of the same comprehension questions for each. However, four questions were appropriate for all three narratives. These questions pertained to the topic of the story; the ending of the story; an alternative ending for the story; and what might happen following the event(s) described in the story. The six remaining questions were unique to the individual narrative passages. As appropriate for each narrative, these questions focused on information such as the time and place of the events in the story, the reason for the events of the story; and problems or obstacles encountered in the story. It is important to note that these questions differed from the ones used by Hustad and Beukelman (2002). The original questions developed by Hustad and Beukelman were designed to query sentence-level information so that each question could be answered on sole basis of its referent sentences. In the present study, questions were designed to query narrative-level information that was not tied exclusively to one sentence.

In the present study, all comprehension questions for each of the three paragraphs were pilot tested to meet two criteria. The first criterion required that questions were not “guessable” to judges who *had not* been exposed to the target narrative. This was established by having a pool of 10 independent judges who had no knowledge of the referent narratives provide answers to the comprehension questions. Questions that were answered correctly by more than 1 of 10 judges were discarded. Questions that were answered incorrectly by at least 9 of 10 judges were deemed “unguessable” and were maintained in the pool of questions.

The second criterion for comprehension questions required that they were “answerable” to judges who *had* been exposed to the target narrative. This was established by having a separate pool of 10 independent judges provide answers to the comprehension questions following presentation of the referent narratives. Questions that were answered correctly by at least 9 of 10 judges were maintained in the pool of questions; and those that were answered incorrectly by more than 1 of 10 judges were discarded. Several iterations and modifications were required before a series of 10 questions for each narrative, each of which targeted a different response, met the two criteria with 10 different judges. The questions that were ultimately selected following all pilot testing were not guessed correctly by any of the 10 naïve independent judges; and were answered correctly by all of the judges who had been exposed to the referent narrative. See Table 2 for a sample narrative, comprehension questions, and one sample listener’s responses to the comprehension questions.

Procedures

Recording speech samples—Each of the 12 speakers was recorded on digital audio tape (48 kHz sampling rate; 16-bit quantization) while producing 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. See Hustad (2006) for details regarding recording procedures.

Preparing speech samples for playback to listeners—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. Stimulus files for each sentence were

normalized using Sound Forge 4.0 (Sonic Foundry, 1989) so that the peak amplitude of each sentence was constant across all files.

Experimental task—Listeners completed the experiment independently in a sound-proof 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.

All experimental tasks were presented via computer using Microsoft Powerpoint. Tasks were self-paced, so that listeners were able to advance through the experiment at a comfortable rate, taking as much time to generate responses as necessary. For all tasks, listeners initiated presentation of the speech sample by clicking the mouse. They were able to hear each sample only one time. For the comprehension task, the entire narrative was presented continuously, with a brief pause between each sentence. Following presentation of the narrative, 10 randomly ordered comprehension questions were presented one at a time, each on a separate screen. Listeners typed their responses to comprehension questions into the computer on the same screen upon which the individual question was presented. They were unable to refer to their responses for previous questions or to view forthcoming questions.

For the intelligibility task, individual sentences forming the narrative were presented individually. Following presentation of one sentence, listeners made orthographic transcriptions of what they heard by typing on the computer. When they were finished typing their response, they proceeded to the next sentence until they had entered transcriptions for all 10 sentences.

Prior to beginning the experimental tasks, listeners were instructed that they would be listening to one person who has a speech disability and that they would complete two types of tasks—comprehension and intelligibility. In the comprehension task, listeners were told that they would hear a story comprised of 10 sentences and then they would answer questions about what they heard after the story was presented in its entirety. In the intelligibility task, listeners were told that they would hear a different story comprised of 10 sentences. Between each sentence, they would be asked to type exactly what they thought the speaker said. The listeners were instructed to type all words that they could, guessing at words if necessary. They were told to skip any words for which they were unable to venture a guess. Listeners were instructed to follow all directions and answer all questions presented on the computer screen. They were also informed that the person speaking would be difficult to understand and that if they were uncertain they should take their best guess.

Randomization and counterbalancing—In each of the two experimental tasks (comprehension and intelligibility) completed by listeners, different narrative passages were employed. To prevent an order effect, presentation of tasks was counterbalanced within each of the 12 speaker groups so that half of the listeners completed the comprehension task first and half of the listeners completed the intelligibility task first. In addition, for the comprehension task, all comprehension questions were presented in random order for each listener, so that no two listeners received comprehension questions in the same order. Finally, speech stimuli associated with the comprehension task and the intelligibility task were also counterbalanced so narratives were represented the same number of times in each condition.

Dependent variables

Two different dependent variables were of interest for this study. The first variable was intelligibility of individual words comprising the narrative produced by the speakers with dysarthria. These data were used to characterize how well listeners processed the surface code of the narratives produced by the speakers. The second variable was experimenter ratings of the accuracy of listener responses to comprehension questions pertaining to the narrative as a whole. These data were used to characterize how well listeners comprehended the situation and events described in the narrative.

Scoring—Orthographic transcriptions generated by listeners were scored on a sentence by sentence basis. Within each transcribed sentence, individual transcribed words were compared with the words produced by the speaker to determine whether there was an exact phonemic match, without regard for word order. Mis-spellings and homonyms were accepted as correct. Each correct word earned 1-point. Across each of the 10 sentences 65 points were possible, one for each target word. The total number of words identified correctly was divided by the total number of words possible and multiplied by 100 to yield the percent of words identified correctly for each listener.

Comprehension of the narratives was determined by scoring listener responses to comprehension questions. The author scored all comprehension questions using a three-point scale (0, 1, 2). Responses to comprehension questions that were judged to be incorrect earned 0-points. Responses that were judged to be general and non-specific, yet not incorrect, earned 1-point. Responses that were judged to be specific and correct earned 2-points. Across each of the 10 questions per narrative, a total of 20 points was possible. Comprehension scores were converted to a percent for each listener by dividing the number of points earned by the number of points possible and multiplying by 100.

Reliability—Both intra- and inter-judge reliability were determined for the percent of words transcribed correctly (intelligibility), and ratings of listener responses to comprehension questions (comprehension). Intra-judge reliability was obtained by having the original judge (the author) re-score data from three randomly selected listeners for each of the 12 speaker-groups (25% of the sample). Inter-scorer reliability was obtained by having a second judge (research assistant), who was not involved in the initial scoring, score data from three randomly selected listeners for each of the 12 speaker-groups (25% of the sample). Unit-by-unit agreement (Hegde, 1994) was obtained by dividing the number of agreements by the number of agreements plus disagreements, multiplied by 100. Cohen's Kappa (Siegel & Castellan, 1988), a measure that accounts for the proportion of inter- or intra-judge agreement expected based upon chance alone, was also calculated. Kappa-type statistics that are .81 and above have been regarded as "almost perfect" (Landis & Koch, 1977).

For intelligibility, both intra-judge and inter-judge unit-by-unit agreement were 100%. Cohen's Kappa was 1.0 for inter- and intra-judge reliability. This is not surprising given that the task of determining whether or not a word is correct is a relatively simple one.

For comprehension, intra-judge point-by-point agreement was 92%, and Cohen's Kappa was .82. Inter-judge point-by-point agreement was 90%, and Cohen's Kappa was .81. These findings document strong intra- and inter-judge reliability in the scoring of dependent measures.

Experimental Design and Statistical Procedures

A 2 × 4 split plot design (Kirk, 1995) was employed for this study. The within subjects variable was 'Measure' and its two categories were intelligibility and comprehension. The between subjects variable was 'Severity' and its four groups were mild, moderate, severe, and profound. Three speakers comprised each severity group; and 12 listeners heard each speaker, for a total of 36 different listeners per severity group.

The research questions of interest pertained to relationships among measures at different levels (across all speakers; within severity groups; within individual speakers), therefore, three sets of analyses were completed using Pearson product moment correlation coefficients. The amount of variance accounted for by the relationship between the two variables (r^2) was of primary interest. Any correlation coefficient that had a probability of .05 or less was considered statistically significant.

Because the two measures, intelligibility and comprehension, addressed different constructs and had different measurement scales, statistical comparisons of mean differences were not made. However, descriptive results are presented and discussed.

Results

Descriptive statistics summarizing mean intelligibility and mean comprehension scores suggest that comprehension scores were consistently higher than intelligibility scores. This was the case across all speakers and listeners, within each of the severity groups, and within each individual speaker. It is noteworthy, however, that variances at all levels (overall, within severity groups, within speakers) were quite large for both intelligibility and comprehension data, suggesting marked variability in performance among listeners. Severity group means for each measure are shown in Figure 1; and individual speaker means are shown in Figure 2.

Inferential statistics, provided in Table 3, showed that when the variability associated with severity group was partialled out, the correlation between intelligibility and comprehension residual scores across all speakers and listeners was .056. This was not statistically significant.

Within severity groups, inferential statistics showed that the relationship between comprehension and intelligibility was significant only for listeners who heard the mild speakers ($r = .341$; $p = .042$). This relationship was not significant for listeners of moderate, severe, or profound speakers. See Figure 3 for scatter plots by speaker severity group.

Within individual speaker groups, inferential statistics showed that the relationship between comprehension and intelligibility was significant for listeners of only two speakers. For listeners of Speaker 8 (mild dysarthria), the correlation was .598 ($p = .040$); and for listeners of Speaker 10 (mild dysarthria) the correlation was .698 ($p = .012$). This relationship was not significant for listeners of the remaining 10 speakers.

Discussion

The purpose of the present study was to evaluate the relationship between comprehension and intelligibility for listeners of speakers with dysarthria within four different severity groups. This study examined the relationship between intelligibility and comprehension scores across all participants when severity effects were statistically controlled, the relationship between intelligibility and comprehension within each severity group, and the

relationship between intelligibility and comprehension for individual speakers with dysarthria. Findings are discussed in detail below.

Relationships between intelligibility and comprehension

Across all speakers and their listeners, results of this study showed that there was no significant relationship between intelligibility scores and comprehension scores when severity effects were removed. This finding suggests that intelligibility scores and comprehension scores reflect different underlying phenomena that do not seem to overlap in meaningful ways when the effects of severity are controlled. From a clinical perspective, this finding has important implications. Perhaps most noteworthy is that generalization of an intelligibility score to conclusions about listeners' ability to comprehend a speaker would likely be inaccurate. Indeed, examination of descriptive data (see Figures 1 and 2) suggests that comprehension scores tended to be higher than intelligibility scores, particularly for listeners of speakers in the moderate and severe groups.

One reason for the descriptive differences and poor relationship between intelligibility scores and comprehension scores in the present study may relate to listeners' goals and their subsequent approach to the two tasks. In the comprehension task, it is likely that listeners were focused on constructing a coherent global picture, actively drawing upon their world knowledge as the narrative unfolded so that they could answer questions about what they heard. In the transcription task it is likely that listeners were focused on lexical delimitation, perhaps with less regard for meaning than in the comprehension task.

Another difference between the measures that may explain the poor relationship and descriptive differences between intelligibility and comprehension scores relates to short term memory limitations of listeners. In the intelligibility task, listeners transcribed sentences one-at-a-time. Listeners were unable to enter their transcription for individual sentences until production of that sentence was complete. Speech rate was markedly reduced relative to normal for most of the speakers, with some sentences lasting as long as 7 or 8 seconds. It is possible that listeners performed worse on intelligibility tasks because their short term memory was taxed due to reduced rate of speech along with increased processing demands imposed by the dysarthric speech signal. Because successful decoding of and short term memory for exact word strings was relatively less important for the comprehension task, the same phenomenon likely had little or no influence on comprehension scores.

Although overall findings of this study showed that there was no relationship between comprehension and intelligibility scores, within speaker severity groups and within individual speaker groups there were some exceptions that warrant discussion. For example, the relationship between intelligibility and comprehension was not significant for speakers and their listeners within the moderate, severe, and profound severity groups. However, for speakers and their listeners within the mild severity group, the relationship between intelligibility and comprehension was significant, although weak (12% of the variance accounted for by the relationship between the two variables). One explanation for this finding is that the speech signal was good enough to allow listeners to understand most of the words produced by speakers; consequently, listeners were able to comprehend the narratives with a high degree of accuracy and the relationship between these measures was significant. That the relationship between comprehension and intelligibility scores was not stronger is somewhat surprising for these speakers and their listeners. In fact, examination of data for individual speakers and their listeners shows that the correlation between intelligibility and comprehension was significant for only one of the three speakers with mild dysarthria. Oddly, this speaker had the lowest intelligibility and comprehension scores within the mild group.

Only one other speaker showed a significant relationship between intelligibility and comprehension scores. This speaker was in the moderate group and did not have any apparent characteristics that distinguished him from the other speakers. Thus the relationship between intelligibility and comprehension for this individual is, again, difficult to explain.

Results of the present study were consistent with those of Hustad and Beukelman (2002) for speakers with severe dysarthria. The replication of findings is particularly noteworthy because there were some important methodological differences between the two studies. In the present study, listeners heard stimuli produced by speakers with dysarthria only one time. In the Hustad and Beukelman study, listeners heard the narrative two times. The effect of hearing target narratives twice may have served to increase intelligibility scores because listeners had the advantage of processing the entire narrative at multiple levels of representation prior to transcribing what they heard.

Another methodological difference involved the scoring rubric employed for comprehension questions. In the present study a three-point scoring system was adopted in which responses that were partially correct were given partial credit. In the Hustad and Beukelman (2002) study, a binomial (2-point) scoring system was employed so that responses to comprehension questions were considered either correct or incorrect. This may have inflated comprehension scores in the present study to some extent. It is also worth noting that use of a three point scale to score comprehension responses is probably more ecologically valid than a binomial scale as comprehension is not necessarily an all or none phenomenon.

Limitations

There were several important limitations to this study that reduce its external validity. First, the study was experimental in nature; and consequently many variables were carefully controlled. For example, speakers were similar in that they all had cerebral palsy. Research suggests that listeners may perform differently on perceptual tasks such as those used in the present study when presented with dysarthria of different etiology (Klasner & Yorkston, 2005; Liss, Spitzer, Caviness, Adler, & Edwards, 2000). All speakers produced narratives in a sentence-by-sentence fashion following a model. Thus, there were no language formulation requirements that typically co-occur with the task of speaking to communicate something. In addition, all sentences comprising the narratives were grammatically correct and complete, which is not always the case when speakers spontaneously generate narratives.

Listeners were relatively homogeneous in this study. As a group, they were young and educated, with normal hearing and minimal experience communicating with speakers who had dysarthria. Speakers and their listeners were not engaged in a real communication task in this study. Listeners simply heard speakers producing the target narratives and then answered questions and transcribed what they heard in two discrete tasks. This situation is unlike real communication when a speaker is talking to a partner and the partner must respond in some way to the speaker.

Another potential limitation relates to the comprehension task. To measure comprehension, listeners answered questions regarding the content of the narrative that they heard. This task required listeners to give deliberate thought to aspects of the narrative that they may not otherwise have considered. In addition, the post-perceptual time spent reflecting on comprehension questions, brief as it may have been, was not likely consistent with what listeners actually do when trying to comprehend spoken messages. Thus, results from the present study may be different from those obtained using different methodologies to measure comprehension. Development of clinically useable measurement tools to characterize the information-bearing capability of dysarthric speech should be considered.

Theoretical and Clinical Implications

The process by which listeners derive the intended meaning from a speech signal is complex and multi-faceted. Although there is considerable debate in the literature, discourse psychologists agree that input is represented at several levels that interact in various ways between initial perception and comprehension of a message. Results of the present study demonstrated that representations of dysarthric speech at one level of processing (surface code) are not closely related to representations at a higher level of processing (propositional content, situation model) (following van Dijk & Kintsch, 1983). Thus, intelligibility measures do not seem to be a good indicator of how well listeners are able to comprehend the intended meaning of a speaker's message. These findings suggest that a comprehensive theory of the impact of dysarthria on communication abilities must move beyond measures of speech intelligibility to address listener processing at multiple levels of representation.

There are important clinical implications for the mismatch between intelligibility and listener comprehension. One implication relates to communicative functioning, a construct that in fact, may be quite different from intelligibility. A speaker may sound very impaired and listeners may have difficulty transcribing sentences produced by that speaker; but, in situations where contextual cues and world knowledge are available, the information bearing capability (i.e. listener's ability to comprehend the message) of that same speech signal may be adequate for the exchange of meaning. Optimal characterization of dysarthric speech should incorporate multiple indices of speech that are specific to the purpose of the measurement. For example, if a clinician wishes to describe the integrity of the speech signal from an acoustic or surface level perspective, traditional intelligibility measures may be appropriate. However, if a clinician wishes to describe the information-bearing capability of that same speech signal, higher level measures that allow for evaluation of listener comprehension should be employed. Ultimately, the simultaneous use of multiple measures to describe dysarthric speech will permit the development of appropriate compensatory interventions that target the transfer of meaning.

Acknowledgments

The author thanks Katie Rentschler, Julie Auker, Gwen Gottardy, Heidi Ake, Becky Lang, and Rhonda Davis for assistance with development and pilot testing of the comprehension questions and for assistance with data collection from listeners. In addition, the author thanks Caitlin Dardis and Lisa Igl for assistance with reliability analyses. This research was funded, in part, by grant R03 DC005536 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

References

Altmann GT. The language machine: Psycholinguistics in review. British Journal of Psychology 2001;92:129–170.

Beukelman DR, Yorkston K. The relationship between information transfer and speech intelligibility of dysarthric speakers. Journal of Communication Disorders 1979;12:189–196. [PubMed: 438358]

Butcher, KR.; Kintsch, W. Text comprehension and discourse processing. In: Healy, AF.; Proctor, RW., editors. Handbook of psychology: Experimental psychology. Vol. Vol. 4. John Wiley & Sons, Inc.; New York: 2003. p. 575-595.

Duffy, J. Motor Speech Disorders: Substrates, Differential Diagnosis, and Management. 2nd Ed. Elsevier Mosby; St. Louis: 2005.

Foltz, P. Quantitative cognitive models of text and discourse processing. In: Graesser, AC.; Gernsbacher, MA., editors. Handbook of discourse processes. Lawrence Erlbaum Associates; Mahwah, NJ: 2003. p. 487-523.

Garcia J, Cannito M. Influence of verbal and nonverbal contexts on the sentence intelligibility of a speaker with dysarthria. Journal of Speech and Hearing Research 1996;39:750–760. [PubMed: 8844555]

Giolas T, Epstein A. Comparative intelligibility of word lists and continuous discourse. Journal of Speech and Hearing Research 1963;6:349–358. [PubMed: 14071872]

Graesser AC, Millis KK, Zwaan RA. Discourse comprehension. Annual Review of Psychology 1997;48:163–189.

Higginbotham DJ, Drazek AL, Kowarsky K, Scally C, Segal E. Discourse comprehension of synthetic speech delivered at normal and slow presentation rates. Augmentative and Alternative Communication 1994;10(3):191–202.

Hunter L, Pring T, Martin S. The use of strategies to increase speech intelligibility in cerebral palsy: An experimental evaluation. British Journal of Disorders of Communication 1991;26:163–174. [PubMed: 1777397]

Hustad KC. A closer look at transcription intelligibility for speakers with dysarthria: Evaluation of scoring paradigms and linguistic errors made by listeners. American Journal of Speech Language Pathology. 2006

Hustad K, Auker J, Natale N, Carlson R. Improving intelligibility of speakers with profound dysarthria and cerebral palsy. Augmentative and Alternative Communication 2003;19:187–198.

Hustad KC, Beukelman DR. Effects of linguistic cues and stimulus cohesion on intelligibility of severely dysarthric speech. Journal of Speech, Language, and Hearing Research 2001;44:497–510.

Hustad KC, Beukelman DR. Listener comprehension of severely dysarthric speech: effects of linguistic cues and stimulus cohesion. Journal of Speech, Language, and Hearing Research 2002;45:545–558.

Hustad KC, Cahill MA. Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. American Journal of Speech-Language Pathology 2003;12:001–011.

Hustad KC, Jones T, Daily S. Implementing speech supplementation strategies: Effects on intelligibility and speech rate of individuals with chronic severe dysarthria. Journal of Speech, Language, and Hearing Research 2003;46:462–474.

Kent R, Weismer G, Kent J, Rosenbek J. Toward phonetic intelligibility testing in dysarthria. Journal of Speech and Hearing Disorders 1989;54:482–499. [PubMed: 2811329]

Kintsch, W. A cognitive architecture for comprehension. In: Pick, HJ.; van den Broek, PW., editors. Cognition: conceptual and methodological issues. American Psychological Association; Washington, DC: 1992. p. 143-163.

Kirk, R. Experimental design: Procedures for the behavioral sciences. 3rd Ed. Brooks / Cole Publishing; Pacific Grove, CA: 1995.

Klasner ER, Yorkston K. Speech intelligibility in ALS and HD dysarthria: The everyday listener's perspective. Journal of Medical Speech-Language Pathology 2005;13:127–139.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174. [PubMed: 843571]

Liss JM, Spitzer SM, Caviness JN, Adler C, Edwards BW. Lexical boundary error analysis in hypokinetic and ataxic dysarthria. Journal of the Acoustical Society of America 2000;107:3415–3424. [PubMed: 10875386]

Singer, M. Psycholinguistics: Discourse comprehension. In: Kazdin, AE., editor. Encyclopedia of psychology. Vol. Vol. 6. American Psychological Association; Washington, DC: 2000. p. 269-372.

Sound Forge 4.5 [computer software]. Sonic Foundry; Madison, WI: 1998.

Tikofsky RS, Tikofsky RP. Intelligibility as a measure of dysarthric speech. Journal of Speech and Hearing Research 1964;7:325–333. [PubMed: 14239011]

Tjaden KK, Liss JM. The role of listener familiarity in the perception of dysarthric speech. Clinical Linguistics and Phonetics 1995;9(2):139–154.

van Dijk, T.; Kintsch, W. Strategies of Discourse Comprehension. Academic Press; New York: 1983.

Weismer, G.; Martin, R. Acoustic and perceptual approaches to the study of intelligibility. In: Kent, R., editor. Intelligibility in Speech Disorders. John Benjamins Publishing Co.; Philadelphia: 1992. p. 67-118.

Yorkston, KM.; Beukelman, DR. Assessment of Intelligibility of Dysarthric Speech. C.C. Publications; Tigard, OR: 1981.

Yorkston, KM.; Beukelman, DR.; Strand, EA.; Bell, KR. Management of motor speech disorders in children and adults. 2nd Ed. Pro Ed; Austin, TX: 1999.

Yorkston, K.; Beukelman, D.; Tice, R. Sentence Intelligibility Test for Macintosh. Communication Disorders Software; Lincoln, NE: 1996. Distributed by Tice Technology Services, Lincoln, NE

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

Zwaan, RA.; Singer, M. Text Comprehension. In: Graesser, AC.; Gernsbacher, MA., editors. Handbook of discourse processes. Mahwah, NJ: 2003. p. 83-121.

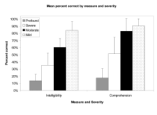

Figure 1.
Mean percent correct by comprehension and intelligibility score and severity group (+SD). Note that raw data were converted to percent correct scores to allow display of comprehension and intelligibility data on the same scale. Intelligibility data are based on 65 possible words correct; comprehension data are based on 20 possible comprehension points.

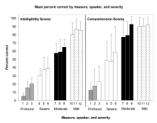

Figure 2.
Mean percent correct by comprehension and intelligibility score, severity group, and individual speaker (+SD). Note that raw data were converted to percent correct scores to allow display of comprehension and intelligibility data on the same scale. Intelligibility data are based on 65 possible words correct; comprehension data are based on 20 possible comprehension points.

Figure 3.
Relationship between listener comprehension and intelligibility scores by severity group. Data plotted are raw percent correct scores ranging from 0 to 100 for each measure.

Table 1

Speaker characteristics.

Speaker	Age	Gender	Dysarthria diagnosis	Dysarthria severity	Perceptual features of connected speech	SIT score	Speech rate in wpm
1	37	M	Spastic	Profound	• Harsh vocal quality, imprecise articulation, short phrases	5%	54
2	33	M	Hyperkinetic-spastic	Profound	• Irregular articulatory breakdowns, imprecise articulation	7%	63
3	24	F	Hyperkinetic-spastic	Profound	• Irregular articulatory breakdowns, imprecise articulation	16%	23
4	58	F	Spastic	Severe	• Hypernasality, imprecise articulation	34%	75
5	46	F	Spastic	Severe	• Harsh vocal quality, hypernasality, imprecise articulation	25%	38
6	42	F	Spastic	Severe	• Hypernasality, imprecise articulation	27%	81
7	21	M	Hyperkinetic-spastic	Moderate	• Harsh vocal quality, irregular articulatory breakdowns, imprecise articulation	48%	56
8	33	F	Spastic	Moderate	• Breathy voice, short phrases, imprecise articulation	50%	102
9	55	M	Spastic	Moderate	• Hypernasality, imprecise articulation, short phrases	70%	84
10	37	M	Spastic	Mild	• Harsh vocal quality, imprecise articulation, short phrases	76%	59
11	32	F	Spastic	Mild	• Breathy voice, imprecise articulation, short phrases	83%	148
12	53	F	Hyperkinetic-spastic	Mild	• Hypernasality, imprecise articulation, short phrases	85%	70

J Speech Lang Hear Res. Author manuscript; available in PMC 2011 January 5.

Table 2

Sample narrative, associated comprehension questions, and responses from one listener who heard Speaker 7. Score of 0 = incorrect response; Score of 1 = vague responses, but not incorrect; Score of 2 = specific and correct response.

Target narrative	Comprehension question	Listener response	Score
Jeffery and Jacob are sports fanatics. College football is their favorite sport. Each year they attend one football game. This year they went to the homecoming game. Kickoff was scheduled for noon. They had a tailgate party first. Their seats were near the fifty yard line. The game went into overtime. The home team won by one touchdown. They wanted season tickets for next year.	• What is another way the story could have ended?	• It could have been a draw	2
	• Based on the outcome of the story, what might happen next?	• They will attend a football game next year.	0
	• How often are the characters involved in the event described in the story?	• Every week	0
	• When did the story take place?	• Didn't specify	0
	• When did the main event described in the story begin?	• Football season	1
	• What is the topic of the story?	• Two boys watching football	2
	• What is one thing, not specifically mentioned in the story, that the characters might wear to the event described in the story?	• Football team attire	2
	• Why are the characters involved in the event described in the story?	• Because they are interested in the sport	2
	• What is the role that the characters play in the event described in the story?	• Spectators	2
	• What was the final outcome of the story?	• One of the teams won by one touchdown	1
		Total points	12
		% correct	60

Table 3

Pearson product moment correlation coefficients for intelligibility and comprehension data.

Source	Comprehension and Intelligibility			
	r	r^2	Observed p-value	
All speakers	.056	.003	.501	
Within Severity Group				
Mild	.341	.116	.042	*
Moderate	−.137	.018	.425	
Severe	.049	.002	.428	
Profound	.136	.018	.777	
Within Individual Speaker Group				
Speaker 1	.091	.008	.778	
Speaker 2	.492	.242	.104	
Speaker 3	−.397	.158	.202	
Speaker 4	.000	.000	1.000	
Speaker 5	.567	.321	.054	
Speaker 6	−.223	.049	.487	
Speaker 7	−.459	.211	.134	
Speaker 8	.598	.358	.040	*
Speaker 9	−.462	.213	.131	
Speaker 10	.698	.487	.012	*
Speaker 11	.347	.120	.269	
Speaker 12	−.384	.147	.217	

*Statistically significant p<.05