Published in final edited form as:

J Speech Lang Hear Res. 2010 October ; 53(5): 1075–1086. doi:10.1044/1092-4388(2010/09-0250).

Relationships between Speech Intelligibility and Word Articulation Scores in Children with Hearing Loss

David J. Ertmer

Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN, US

Abstract

Purpose—This investigation sought to determine whether scores from a commonly used word-based articulation test are closely associated with speech intelligibility in children with hearing loss. If the scores are closely related, articulation testing results might be used to estimate intelligibility. If not, the importance of direct assessment of intelligibility would be reinforced.

Methods—Forty-four children with hearing losses produced words from the Goldman-Fristoe Test of Articulation-2 and sets of 10 short sentences. Correlation analyses were conducted between scores for seven word-based predictor variables and percent-intelligible scores derived from listener judgments of stimulus sentences.

Results—Six of seven predictor variables were significantly correlated with percent-intelligible scores. However, regression analysis revealed that no single predictor variable or multi-variable model accounted for more than 25% of the variability in intelligibility scores.

Implications—The findings confirm the importance of assessing connected speech intelligibility directly.

Keywords

CHILDREN; SPEECH INTELLIGIBILITY; HEARING LOSS

Speech intelligibility has been defined as “that aspect of speech-language output that allows a listener to understand what a speaker is saying” (Nicolosi, Harryman, & Kresheck, 1996). Research completed during the 1960s – 1980s revealed that, despite amplification and speech training efforts, the speech of children with severe to profound hearing loss was approximately 20% intelligible on average (see Osberger, 1992 for review). Clearly, such a low level of intelligibility can lead to substantial communication difficulties at home, in school, and in social settings.

The recent implementation of newborn hearing screening and advances in cochlear implant and hearing aid technology have increased optimism that children with hearing loss can become readily intelligible talkers (e.g., Chin, Tsai, & Gao, 2003; Connor, Craig, Raudenbush, Heavner, & Zwolan, 2006; Ertmer, Young, & Nathani, 2007; and Moeller, Hoover, Putman, et al, 2007). For this outcome to be realized, however, it is essential that speech intelligibility be a closely monitored intervention priority. This investigation examined relationships between scores from a commonly used articulation test and percent-intelligible scores from a sentence imitation task to determine how closely the former scores predicted speech intelligibility levels in children with hearing loss.

Clinical Uses of Speech Intelligibility Measures

Measures of speech intelligibility have at least three vital clinical uses for children with hearing loss. First, speech intelligibility and speech perception scores have been found to be significantly correlated in children with cochlear implants (CIs; Svirsky, Robbins, Kirk, Pisoni, & Miyamoto, 2000). Thus, the extent to which speech production improves is likely to be influenced by the auditory benefits received from sensory aids. Whereas improved intelligibility provides supporting evidence of sensory aid benefit, failure to make gains over time--when considered with other measures of speech perception ability--may indicate limited benefit from hearing aids or cochlear implants.

Intelligibility testing also provides essential information for enrollment/dismissal decisions and intervention planning. Comprehensive speech evaluations address both the phoneme-production *and* connected speech abilities of children. The value of articulation testing in determining the need for intervention and in selecting intervention targets has long been recognized. However, word-based testing reveals little about the clarity of children's connected speech. Intelligibility testing can supply this missing information. Using both types of assessments enables clinicians to make enrollment decisions that are based on phoneme production *and* functional speech abilities, and to identify goals in both of these areas. Intelligibility testing can also provide evidence for determining whether children have developed understandable speech and are ready for dismissal from their programs.

Finally, measures of speech intelligibility can also inform decisions about academic placement. Monsen (1981) concluded that listeners are "confronted by overwhelming difficulty in understanding" when children's speech is less than 59% intelligible (p. 350). This "cut-off" level can serve as a benchmark for estimating how well children's speech will be understood by peers and teachers in regular classrooms. That is, children whose scores are above 59% should experience less difficulty being understood than those below this level. Thus, intelligibility scores, when considered with measures of academic, social, and language ability, enable educators to be more thorough in evaluating readiness for mainstreaming.

Methods of Assessing Speech Intelligibility

There are two main methods of assessing speech intelligibility directly: scaling and item identification (See Kent, Weismer, Kent, & Rosenbek, 1989; Osberger, 1992 for detailed discussion). Scaling involves asking listeners to select a rating within a continuum of intelligibility. For example, children's speech samples can be evaluated using a numeric scale (e.g., 1 – 10 point scale with "1" equated with totally unintelligible and "10" equated with readily intelligible speech) or descriptors (e.g., "not at all," "seldom," "sometimes," "most of the time," and "always") to indicate how well speech is understood. The main advantage of scaling is that it is relatively easy to complete. Clinicians simply record a speech sample and play it for a listener to rate. The main disadvantages of this approach include the subjective nature of the ratings (i.e., different raters may have different internal rating criteria) and the fact that scaling may not distinguish among talkers whose intelligibility falls in the middle range of ability (e.g., 30 – 70% intelligible; Samar & Metz, 1988). The latter limitation makes it difficult to measure progress during intervention. A more objective and sensitive evaluation can be made with item identification procedures.

In item identification tasks, words or sentences are presented to listeners under open-set or closed-set conditions. In the open-set condition, clinicians make recordings of children's speech (i.e., words or sentences) and play them for listeners who write down the words they perceive. The listener's written responses are then scored for the number of words that

match the words the child attempted to say. A percent-intelligible score is calculated by averaging the scores obtained across all of the listeners.

In closed-set tasks, listeners are provided with a set of choices (e.g., pictures or printed word lists) for each child utterance. A percent-correct score is calculated by comparing the listener's choice with the child's intended word. The possibility of selecting the correct word by chance alone must also be considered when interpreting closed-set scores. For example, a score of 25% on a four-choice task does not mean the child's speech is 25% intelligible. The score actually represents the expected level of performance if no words were recognized (i.e., score expected from guessing alone). The Picture Speech Intelligibility Evaluation (Picture SPINE; Monsen, Moog & Geers, 1988) is an example of a closed-set word identification procedure.

Both scaling and item identification outcomes are influenced by factors such as listener familiarity with talkers who have hearing loss, the presence/absence of speechreading cues, knowledge of context, and the number of times speech samples are presented to listeners (see Osberger, 1992, for review). Speech samples are often presented twice in the auditory-only condition. Listeners who have little or no exposure to the speech of deaf children are recruited to obtain a rigorous estimate of intelligibility (see Osberger, 1992 for review).

Implementation of Speech Intelligibility Assessments

Despite vital clinical purposes and readily available procedures, evaluations of speech intelligibility appear to be underused with children who have hearing loss. As Monsen observed more than 25 years ago, "A strange fact about the contemporary education of the hearing-impaired is that the intelligibility of their speech is seldom measured." (1981; p. 845). Based on previous clinical experience and the author's discussions with clinicians, supervisors, and administrators at several schools for deaf children, Monsen's comment appears to have face validity today as well.

Several factors appear to contribute to the limited use of direct assessments of speech intelligibility in educational settings. First, speech-language pathologists often rely on their own observations of spontaneous speech and easy-to-administer articulation test results to estimate connected speech proficiency. There are at least two potential problems with this approach. First, subjectivity may lead to an overestimation of children's connected speech intelligibility. That is, clinicians who are familiar with a child's speech are likely to understand it more readily than persons who do not know the child (Osberger, 1992). Second, little is known about the relationship between commonly used word-articulation test scores and intelligibility levels in children with hearing loss. Studies by Dubois and Bernthal (1978) and Johnson, Winney, and Pederson (1980) demonstrated that children with normal hearing who had speech disorders produced phonemes with much greater accuracy in single words than in connected speech. A similar pattern in children with hearing loss would lead to overestimation when word-based articulation tests are used to predict intelligibility levels.

Interviews with administrators and clinicians at several schools for deaf children revealed the common perception that scaling and item identification procedures were time-consuming and impractical because extra time is needed to record samples and recruit listener-judges. Thus, the ease of articulation testing and the extra time needed to conduct rating or word identification tasks appear to practical concerns that dissuade clinicians from direct assessment of intelligibility.

Although there are limitations in using articulation tests to estimate intelligibility, it is possible that some word-based scores might have clinical usefulness. That is, if there is a strong relationship between some scores derived from articulation testing (e.g., percent of

initial consonants correct, percent of vowels correct) and speech intelligibility, then a single score or a combination of scores might be used to develop a reference table for estimating the clarity of connected speech.

Purposes of the Study

The purpose of this investigation was to examine the relationships between seven individual phonological scores calculated from the Sounds in Words subtest of the Goldman-Fristoe Test of Articulation-2 (GFTA-2; Goldman & Fristoe, 2000) and children's sentence-level speech intelligibility scores derived from an item identification task. A multivariate regression analysis was also used to determine the degree to which combinations of phonological scores predicted speech intelligibility. If the predictor variables are closely related to intelligibility scores, then they might be useful for estimating children's intelligibility. If not, the need for direct assessment of intelligibility would be reinforced.

Methods

Participants

Forty-four children with bilateral hearing losses participated in the study. The severity of their hearing losses ranged from mild to profound. All used hearing aids or cochlear implants and attended oral education programs in the Midwestern United States. The children ranged in age from 2;10 to 15;5 (years; months) at the time of testing. Table 1 contains background information on gender, age, sensory aids, hearing levels, and the sentence list used to elicit speech samples.

Stimulus Materials and Data Collection

The Goldman-Fristoe Test of Articulation-2; Sounds-in-Words subtest (Goldman & Fristoe, 2000) was administered to each child following standard procedures. The GFTA-2 was selected to elicit spoken words because it is widely used by speech-language pathologists in both regular schools and schools for children with hearing loss and was standardized to represent children of various races, economic levels, and geographical locations. The Sounds-in-Words subtest of the GFTA-2 consists of 34 words represented by color drawings. Children are asked to name each drawing and are given cues and spoken models if they do not respond spontaneously or if they misname items. The acceptability of consonant phonemes is assessed in the initial, medial, and final positions in words. Consonant cluster acceptability is assessed in the initial position only. Vowels are not included in the standard analysis of GFTA-2 results. However, because vowel production errors are common in children with hearing loss (Levitt & Stromberg, 1983; Monsen, 1976, 1978), vowel acceptability was examined in the current investigation.

Samples of children's connected speech were gathered using two sets of sentences. The Beginner's Intelligibility Test (BIT; Osberger, Robbins, Todd, & Riley, 1994) was given to young children who could not read. There are four BIT lists and each list consists of 10 grammatically simple sentences incorporating vocabulary likely to be familiar to young children (see Appendix A). Testers modeled each sentence while they demonstrated its meaning with objects and toys. For example, for "The boy walked to the table," the tester moved a small boy figure toward a toy table while saying the sentence. Children were asked to repeat each BIT sentence after the tester. The Monsen – Indiana University sentences (M-IU sentences; Osberger, Robbins, Todd, & Riley, 1994) were given to children in second grade or higher who were able to read. It consists of three lists of 10 sentences (see Appendix B). The M-IU sentences differed from the BIT sentences in that they were slightly longer and included more consonant clusters. Each M-IU sentence was printed on an index

card and shown to children. Children listened each sentence and then repeated the sentence after the index card was turned over so that the influence of reading on speech was lessened.

All speech samples were recorded using Marantz audio cassette recorders coupled with lapel microphones placed within 6 inches of the mouth. The samples were prepared for presentation to listeners by isolating each sentence from the tester's models, digitizing the children's sentences at a sampling rate of 20 kHz, and storing them as a computer files. Each child's digitized sentences were made into a playlist using Roxio software, and burned to compact discs. Each play-list consisted of an announcement of the number of each sentence followed by two presentations of that sentence.

Data Analysis

Transcription—Words produced during the administration of the GFTA-2 were broadly transcribed using the International Phonetic Alphabet by undergraduate and graduate students in the department of Speech, Language, and Hearing Sciences at Purdue University. All of the transcribers passed hearing screenings and had completed coursework in Phonetics. Singleton consonant and consonant cluster targets were judged as acceptable or unacceptable (i.e., correct or misarticulated productions) for target consonants in the initial, medial, and final positions in words. Vowel acceptability was assessed for all vowels in GFTA-2 target words. A percent correct score was obtained for seven predictor variables: Total targeted consonants and clusters in all word positions ($n = 77$), Initial consonants without clusters ($n = 22$), Initial consonants including clusters ($n = 38$), Medial consonants ($n = 20$), Final consonants ($n = 19$), Consonant clusters ($n = 16$), and Vowels ($n = 82$).

Intelligibility Measures—Speech intelligibility was measured by having adults listen to audio-recordings of the children's sentences and indicate the words they understood. All listeners had normal hearing and reported little or no exposure to the speech of children with hearing loss. The listeners wore headphones as they listened to the sentences and wrote down what they understood. They were encouraged to guess whenever unsure of a word. They were also asked to write an X for every unintelligible word so that they remained focused on the task even when speech was difficult to understand. Written responses were scored by counting the number of words that were correctly identified by each listener. The total number of correctly identified words was then divided by the total number of words that the child attempted to produce. For example, if a child said "Boy walking" instead of the target sentence "The boy is walking to the table," only two words (rather than seven) were used in calculating the percentage of words correctly identified. Children were given credit for producing a word if the listener identified the root of the word. For example, if a child said "Mommy jump" instead of "Mommy jumps," the word "jump" was counted as correct if identified. Three adults listened to the sentences produced by each child. A percent-intelligible score was calculated for each child by averaging the percent of words identified across the three judges. Adults listened to a specific BIT or M-IU list only once so that prior exposure to a list did not influence subsequent scores.

Reliability—Intra- and inter-transcriber reliability was calculated by having the original transcriber re-transcribe the GFTA-2 for 9 children (20%) and having a second transcriber transcribe the GFTA-2 for 15 children (34%). Both forms of reliability were assessed to determine item-by-item agreement on acceptability. Intra-transcriber agreement was found to be 96% for all target consonants and vowels. Inter-transcriber agreement was 93% for the same targets and vowels. The consistency of listener judgments on the BIT and M-IU sentences was assessed by calculating the average difference between judges' scores for sets of 10 sentences. On average, members of each panel of listeners differed from each other by 4.97 percentage points ($SD = 3.73$).

Statistical Analyses

Raw scores from the GFTA-2 were converted to percent correct scores and Pearson product moment correlations were calculated to determine the strength of their relationship with speech intelligibility scores. Total Target Consonants (TTC), Initial Consonants with consonant clusters (ICWC), Medial Consonants (MC), Final Consonants (FC), Initial Consonants without Clusters (IC), Consonant Clusters (CC), and Vowels and diphthongs (V) were also examined in a linear regression analysis to determine how much of the variability in intelligibility scores was accounted for by each predictor variable. Because it was anticipated that single variables might not account for high levels of the variability in intelligibility scores, a multiple linear regression analysis was completed to identify the two- and three- variable models that had the best predictive value based on adjusted R-square values.

Results

Relationships between Word-based Scores and Speech Intelligibility

Descriptive Data—Intelligibility and percent correct scores for predictor variables can be found in Table 2. The mean intelligibility score across all children was 54.5% (range 0 – 97%). Mean scores for the predictor variables were considerably greater. On average, vowels were produced with the greatest accuracy (97%; range 84.5 – 100%), followed by Final Consonants ($M = 92.3\%$; range 42.1 – 100%), Initial Consonants (without clusters) ($M = 90.6\%$; range 45.5 – 100%), Medial Consonants ($M = 88.4\%$; range 35 – 100%), Total Target Consonants ($M = 86\%$; range 37.6 – 100%), Initial Consonants With Clusters ($M = 82.3$; range 36.8 – 100%), and Consonant Clusters ($M = 71\%$; range 25 – 100%). In summary, despite moderately high accuracy in producing consonants and vowels in single words, only slightly more than half of the words from the BIT and M-IU sentences were understood by listeners.

Correlations for Predictor Variables—Figure 1 contains scatter plots for each of the seven predictor variables and the children's intelligibility scores. Correlation coefficients and probability values are also included in each scatter plot. From this information it can be seen that—except for Final Consonants—each predictor variable was significantly correlated with speech intelligibility, but the association was not especially strong. For instance, the strongest correlation (Initial Consonants without clusters; $r = .51$) accounted for only 26% of the variability in intelligibility scores. Slightly weaker correlations were noted for Total Targets Correct, Initial Consonants with clusters, and Medial Consonants. Vowels and Consonant Clusters had the two weakest correlations with intelligibility, with near-ceiling level scores limiting the relationship between Vowels and intelligibility. Intra-variable correlations (Table 3) revealed that many of the predictor variables were highly associated with each other ($M: r = .68$).

Multiple Regression Analyses—Linear regression analysis results (Table 4) revealed that the Total Target Consonants score was the best single-variable predictor of intelligibility. Predictor variables were also used in a multiple regression analysis with speech intelligibility as the dependent variable. The goal of this analysis was to find the strongest combination of variables for predicting intelligibility levels. It was determined that the combination of Initial Consonants without Clusters and Consonant Cluster scores was the strongest two-variable combination, and that the combination of Total Target Consonants, Final Consonants, and Consonant Cluster scores was the best three-variable model for predicting intelligibility. However, as Table 4 shows, none of these models accounted for more than 25% of the variance in intelligibility scores. Thus, word-based scores proved to be poor estimators of children's sentence-level speech intelligibility.

Discussion

Relationships between Articulation and Intelligibility Scores

Correlations between the selected articulation measures and intelligibility scores were significant, but not especially strong. A close look at the scatter plot for the strongest predictor variable, Initial Consonants without clusters, reveals that scores of 80% or better for this predictor were almost evenly associated with intelligibility scores >60% as well as those <60%. Thus, this variable was not useful in determining whether children had reached the 59% cut-off level that Monsen (1981) proposed as the threshold to listener understanding. Another example of the limitations of the correlations can be seen in Vowel scores that ranged from 90 – 100% correct whereas intelligibility scores ranged from 0 – 100%. The combination of relatively high phonological scores (near ceiling level in many cases) and a wide ranging intelligibility scores is evident in the remaining scatter plots, as well. Thus, *individual* word-based scores were not strong predictors of connected speech intelligibility.

Multiple regression analyses were conducted to determine whether two or three predictor variables could be combined to account for a sizable portion of the variability in intelligibility scores. If this was the case, the combined variables might be used to formulate a two- or three-factor reference table for estimating children's intelligibility. Unfortunately, the best of these models (TTC, FC, and CC) accounted for only approximately 25% of the variability in children's intelligibility scores and would, therefore, have very limited clinical value (Table 4).

In summary, scores from the Sounds in Words subtest of the GFTA-2 indicated that the participants had achieved relatively high, word-level phonological skills. These skills, however, did not transfer into high levels of intelligible connected speech. What might account for the disconnection between these two sets of scores?

Relationships between Word Articulation Scores and Speech Intelligibility

Several factors and their interactions may have contributed to the limited correspondence between children's word articulation skills and their intelligibility levels. To begin with, the target words were shorter and less complex than the target sentences. GFTA-2 words averaged approximately 1.5 syllables and 4.4 phonemes each. In comparison, the sentences of the BIT (4.6 syllables and 11.9 phonemes, respectively) and the M-IU materials (5.5 syllables and 12.7 phonemes, respectively) were considerably longer and more phonetically complex. Sentence production also involves both suprasegmental (i.e., stress, timing, and intonation) and linguistic (syntactic and morphologic) features not necessary for the production of single word production. The combination of greater phonetic, suprasegmental, and linguistic complexity found in sentences likely contributed to lower intelligibility scores.

Potential reasons for differences in phoneme accuracy in spoken words Vs sentences have also been considered for children with normal hearing. Dubois and Bernthal (1978) and Johnson, Winney, and Pederson (1980) found that phonemes were produced with much greater accuracy in single words than in sentences by children with speech disorders. Johnson, et al (1980) proposed that a "more deliberate articulatory focus" (p. 179) during word production—as compared to sentence production—might explain this discrepancy. The authors further cautioned that there are serious limitations in using single word tests to estimate connected speech proficiency in children with speech disorders who have normal hearing. The findings of the current study strongly suggest that this caution applies to children with hearing loss as well.

Speech training practices might also contribute to the discrepancy between word articulation and intelligibility scores. Single words are often practiced before sentences in commonly used intervention programs such as the Traditional approach (Van Riper & Emerick, 1984); Phonological Processes remediation (Hodson & Paden, 1983); and the Ling approach (Ling, 1980). Transfer of phonological learning from words to connected speech requires generalization of skills to more complex speech tasks, self-monitoring, and higher-level understanding on the part of the child (McReynolds, 1981). Children with hearing loss are likely to need practice in producing targets in connected speech if they are to be understood more easily. If practice in producing targets in phrases, sentences, and conversation is neglected, connected speech intelligibility might be slow to improve. Suggestions for optimizing carryover to conversational speech can be found in Ertmer and Ertmer (1998).

Although scores from a word-based articulation test were not found to be useful for predicting intelligibility, scores from sentence-based articulation tests might yield better results. For example, Johnson et al (1980) found more speech errors on the Words in Sentences subtest of the GFTA than the Sounds in Words subtest. Preliminary data with the ALPHA test (ALPHA; Lowe, 2000) also suggests that relatively closer estimates of intelligibility may be possible when carrier sentences are used to elicit speech samples from children with hearing loss (Ertmer, 2007). Further study is needed to determine whether the results of sentence-level tests can account for enough of the variability in intelligibility scores to be clinically useful.

Options for Direct Assessment of Connected Speech Intelligibility

The limited correspondence between word articulation and connected speech scores highlights a need to assess connected speech intelligibility directly. Several simple assessment procedures and modern digital technology can help to make such an evaluation more practical and time-efficient.

Adaptation of Scaling Measures—As mentioned earlier, scaling tasks require listeners to rate children's speech with numbers or descriptive phrases. Scaling involves recording speech samples (e.g. story retelling, a list of sentences) and replaying them for adult listeners. Concerns about the subjective nature of listener's perceptions can be lessened by asking the same listeners to be raters for different children. In this way, the same internal standards are applied to each child. Subjectivity can be reduced further by using unambiguous descriptors such as "never understood," "understood sometimes," "mostly understood," "almost always understood," rather than numeric rating scales with unclear values. Such descriptors, although not capable of identifying mid-range progress or small changes, can give an indication of emerging and high level communicative competence when ratings of "mostly understood" or "almost always understood" are obtained from more than one unfamiliar listener.

Item Identification Measures—Clinicians might also use an item identification task like the one used in the current study. If this approach is chosen, the sentence lists in the appendices can be used to elicit imitative speech samples. The BIT sentences are intended for use with preschool and early elementary children because the sentences are short, have familiar vocabulary, and toys are used to engage the child. The M-IU sentences are appropriate for older students who can read.

The recent development of hand-held digital recorders has made item identification tasks much easier to complete. Now, instead of the complicated and time-consuming procedures used to digitize files and make playlists on personal computers, clinicians can make digital recordings and playlists simultaneously. Using these relatively inexpensive devices, each

sentence is automatically saved as a numbered, digitized file as the recorder is turned on and off, thus eliminating the need to make playlists. The files can then be played for listeners in a rating or item identification task. When a rating procedure is used, adults listen to the entire recording (e.g., a story retelling or a list of sentences) before giving an overall rating. Using an item identification approach, listeners write the words they understood after listening to each sentence twice. Their written responses are then compared to the actual words that the child attempted to say (see scoring conventions in the Methods section; Ertmer, 2008), and a percent-intelligible score is calculated.

Both of these measures require that listeners use headphones to lessen background noise. Different stories and sentence lists should be used each time children are recorded so that listeners do not become familiar with the materials. Listeners who have little exposure to the speech of children with hearing loss provide the most rigorous assessment. Regular assessments (e.g., every 6 months) can be an effective way to document progress toward IEP goals. The methods described above are time-efficient. Stories and sets of 10 sentences can be collected in under five minutes and listeners usually require 10 minutes or less per child.

Conclusions

This investigation clearly demonstrated that word-based articulation test scores do not provide reliable estimates of connected speech intelligibility. It also highlighted --by the lack of close relationships between these scores-- the importance of assessing connected speech intelligibility directly. When such assessments are completed on a routine basis, they can provide vital information for assessing sensory aid benefit, making intervention plans, and determining readiness for regular classroom placements.

Acknowledgments

This study was completed through the support of grants from the National Institutes on Deafness and other Communication Disorders (R01DC007863) and The Purdue University Research Foundation. We are especially indebted to the children and parents who made this study possible. Special thanks to Michele Wilkins, Wendy Adler, and Dawn Violette at Child's Voice School in Wood Dale, IL, and Mary Daniels, Roseanne Siebert (CSJ), and Beverly Fears at the St. Joseph Institute in Chesterfield, MO, and the faculty of these schools for their enthusiastic assistance in carrying out this study. The contributions of Anna Brutsman, Katie Connell, Katie Dobson, Christy Macak, Katie Masterson, Lara Poracki, Elizabeth Robinson, Jennifer Slanker, Elesha Sharp, and Alicia Tam in preparing audio CDs, calculating articulation scores, and estimating reliability are gratefully recognized. Thanks also to Bruce Craig, Yong Wang, and Benjamin Tyner of the Purdue University Statistical Consulting program for their assistance in conducting statistical analyses.

References

- Chin SB, Tsai PL, Gao S. Connected speech intelligibility of children with cochlear implants and children with normal hearing. *American Journal of Speech-Language Pathology*. 2003; 12:440–451. [PubMed: 14658996]
- Connor CM, Craig HK, Raudenbush SW, Heavner K, Zwolan TA. The age at which young deaf children receive cochlear implants and their vocabulary and speech-production growth: Is there added value for early implantation? *Ear and Hearing*. 2006; 27:628–644. [PubMed: 17086075]
- Dubois E, Bernthal J. A comparison of three methods for obtaining articulatory responses. *Journal of Speech and Hearing Disorders*. 1978; 43:295–305. [PubMed: 692096]
- Ertmer, DJ. Speech Intelligibility in Young Cochlear Implant Recipients: Gains and Relationships with Phonological Measures. Poster Presented at the 11th International Symposium on Cochlear Implants in Children; Charlotte, N.C. 2007 Apr.
- Ertmer DJ. Speech intelligibility in young cochlear implant recipients: Gains during year 3. *Volta Review*. 2008; 107:85–99.
- Ertmer DJ, Ertmer PA. Constructivist strategies in phonological intervention: Facilitating self-regulation for carryover. *Language, Speech, and Hearing Services in the Schools*. 1998; 29:67–75.

- Ertmer DJ, Young NM, Nathani S. Profiles of vocal development in young cochlear implant recipients. *Journal of Speech, Language, and Hearing Research*. 2007; 50:393–407.
- Goldman, R.; Fristoe, M. Goldman Fristoe-2 Test of Articulation. Minneapolis, MN: Pearson Assessments; 2000.
- Johnson JP, Winney BL, Pederson OT. Single word versus connected speech articulation testing. *Language, Speech, and Hearing Services in the Schools*. 1980; 11:175–179.
- Kent RD, Weismer G, Kent JF, Rosenbeck JC. Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders*. 1989; 54:482–499. [PubMed: 2811329]
- Levitt, H.; Stromberg, H. Segmented characteristics of the speech of hearing-impaired children: Factors affecting intelligibility. In: Hochberg, I.; Levitt, H.; Osberger, MJ., editors. *Speech of the hearing impaired: Research, training, and personnel preparation*. Baltimore: University Park; 1983. p. 53-74.
- Ling, D. *Speech and the hearing-impaired child: Theory and practice*. Washington D. C: A. G. Bell Association for the Deaf; 1980.
- Ling, D. *Foundations of spoken language for the hearing-impaired child*. Washington D. C: A. G. Bell Association for the Deaf; 1988.
- Lowe, R. *Assessing the Link between Phonology and Articulation -Revised*. Mifflinville, PA: Speech and Language Resources; 2000.
- McReynolds LV. Generalization in articulation training. *Analysis and Intervention in Developmental Disabilities*. 1981; 1:245–258.
- Moeller MP, Hoover B, Putman C, Arbataitis K, Bohnenkamp G, Peterson B, Wood S, Lewis D, Pittman A, Stelmachowicz P. Vocalizations of infants with hearing loss compared with infants with normal hearing: Part I: Phonetic Development. *Ear and Hearing*. 2007; 28:605–627. [PubMed: 17804976]
- Monsen RB. Normal and reduced phonological space: The production of English vowels by deaf adolescents. *Journal of Phonetics*. 1976; 4:189–198.
- Monsen RB. Towards measuring how well hearing impaired children speak. *Journal of Speech and Hearing Research*. 1978; 21:197–219. [PubMed: 703272]
- Monsen RB. A usable test for the speech intelligibility of deaf talkers. *American Annals of the Deaf*. 1981; 126:845–852. [PubMed: 7304383]
- Monsen, RB.; Moog, J.; Geers, A. *Picture Speech Intelligibility Evaluation*. St. Louis, MO: Central Institute for the Deaf; 1988.
- Nicolosi, L.; Harryman, E.; Kresheck, J. *Terminology of Communication Disorders*. 4. Baltimore: Williams and Wilkins; 1996.
- Osberger, MJ. Speech intelligibility in the hearing impaired: Research and clinical implications. In: Kent, RD., editor. *Intelligibility in Speech Disorders*. Philadelphia: John Benjamins Publishing; 1992. p. 233-265.
- Osberger MJ, Robbins A, Todd S, Riley A. Speech intelligibility of children with cochlear implants. *Volta Review*. 1994; 96:169–180.
- Samar V, Metz D. Construct validity of speech intelligibility rating-scale procedures for the hearing-impaired population. *Journal of Speech and Hearing Research*. 1988; 31:307–316. [PubMed: 3172748]
- Svirsky M, Robbins A, Kirk K, Pisoni DB, Miyamoto RT. Language development in profoundly deaf children with cochlear implants. *Psychological Science*. 2000; 11:153–158. [PubMed: 11273423]

Figure 1. Scatter plots for each of seven predictor variables and speech intelligibility scores. Correlation coefficients and p values included for each pair.

Table 1

Demographic and hearing information.

Child	Gender	Age at Testing (months)	Sensory Aid	Device Manufacturer, Model	Aided Thresholds Right Ear	Aided Thresholds Left Ear	BIT/M-IU ²
ADBR	M	156	Bi-Lat CI	Cochlear Americas, Freedom (2)	17	17	M-IU 1
ALHA	F	109	CI	Advanced Bionics, PSP	23		BIT 4
ANHE	M	67	Bi-Lat HA	Widex, Diva(2)	30	27	BIT 1
ASLO	F	135	CI	Cochlear Americas, Freedom	17		M-IU 3
AUMI	F	139	CI	Advanced Bionics, PSP	26		M-IU 3
BECL	M	72	Bi-Lat HA	Phonak, Supero 412(2)	46	45	BIT 1
BEDO	M	97	R: CI, L: HA	Cochlear America's Freedom, Widex P38	30	56	BIT 1
CHNO	M	121	R: HA, L: CI	Miracle Ear, Interpreter; Adv Bionics, Auria	31	20	M-IU 2
CUBR	M	54	Bi-Lat CI	Cochlear Americas, Freedom (2)	27	29	BIT 4
CUJA	M	89	Bi-Lat HA	Widex, Vita SV-38(2)		28	BIT 2
ELVI	F	41	Bi-Lat HA	Phonak, Maxx 211 (2)	29		M-IU 1
GIRU	F	123	CI	Advanced Bionics, PSP	25	24	M-IU 2
GRHA	F	91	R: CI, L: HA	Advanced Bionics Auria; Widex, P38	20	58	BIT 3
ISPR	M	126	R: CI	Advanced Bionics, PSP	27		M-IU 1
ISRO	F	39	Bi-Lat CI	Cochlear Americas, Freedom (2)		40	BIT 2
JAAL	M	114	Bi-Lat CI	Cochlear Americas, Freedom(2)	19	16	BIT 2
JABA	M	88	Bi-Lat CI	Advanced Bionics, Auria (2)	23		M-IU 1
JAHA	M	149	L: CI	Cochlear Americas, 3G		20	M-IU 2
JAST	F	81	CI	Advanced Bionics, Hi-Res	35		M-IU 3
JEHA	F	68	Bi-Lat HA	Phonak, 311 dAZ (2)	71		M-IU 3
JEHA	M	114	CI	Cochlear Americas, Nuc 24		16	BIT 3
JUCA	F	82	R: CI, L: HA	Advanced Bionics, PSP; Widex Senso P38	23	35	BIT 3
KAHA	F	145	Bi-Lat HA	Widex, Senso Diva(2)	32	37	M-IU 1
KYCO	M	155	Bi-Lat HA	Qualitone, Evolution (2)	29	33	M-IU 3
LEBR	M	120	CI	Cochlear Americas, Freedom	23		BIT 4
MABI	M	51	Bi-Lat HA	Oticon, Gaia (2)	20		BIT 1
MASUL	F	89	CI	Cochlear Americas, Freedom	20		M-IU 1
MATA	F	140	CI	Cochlear Americas, 3G	24		M-IU 1
MAWO	M	35	Bi-Lat CI	Cochlear Americas, Freedom(2)	30		BIT 2

Child	Gender	Age at Testing (months)	Sensory Aid	Device Manufacturer, Model	Aided Thresholds Right Ear	Aided Thresholds Left Ear	BIT/M-IU ²
MIBE	F	186	Bi-Lat HA	Oticon, DigiFocus II(2)	33	39	M-IU 2
MICR	M	133	CI	Cochlear Americas, Freedom	18		M-IU 3
MIKI	M	97	CI	Advanced Bionics, Auria	21		BIT 4
MOFE	F	157	R: HA, L: CI	Phonak, Pico Forte; Cochlear Americas, Freedom	85	20	M-IU 1
NABU	F	77	CI	Cochlear Americas, Freedom		15	M-IU 2
NICO	M	39	HA	Widex, Bravissimo	36 [/]	32 [/]	BIT 1
NILI	F	142	CI	Cochlear Americas, Freedom		12	M-IU 1
NOHI	M	128	R: HA, L: CI	Widex, P38; Advanced Bionics, Auria		24	M-IU 1
RIKL	F	34	Bi-Lat HA	Oticon, Syncro (2)	43 [/]	30 [/]	BIT 3
RODA	M	60	HA	Phonak, Aero	80 [/]	78 [/]	M-IU 2
SAGO	F	131	CI	Advanced Bionics, Platinum	26		M-IU 1
SAMA	F	71	Bi-Lat HA	Oticon, Adapto Power (2)	65 [/]	62 [/]	M-IU 2
SYNA	F	67	Bi-Lat CI	Cochlear Americas, Freedom (2)	15	15	M-IU 1
TABE	F	171	Bi-Lat HA	Cochlear Americas, 3G; Freedom	22	24	M-IU 3
ZANA	F	52	Bi-Lat CI	Cochlear Americas, Freedom (2)	19	20	M-IU 1

¹ These values are for unaided three-frequency averages; aided thresholds were not available in school records.

² Beginners Intelligibility Test/Monsen-Indiana University Sentences

Table 2

Percentage of intelligible words in sentences (Intel) and percent correct scores for each predictor variable: Total Targets Correct (TTC), Vowels (V), Initial Consonants (IC), Medial Consonants (MC), Final Consonants (FC), and Initial Consonants Without Clusters (ICWC).
Sorted by intelligibility scores.

Child	Intel	TTC	V	IC	MC	FC	CC	ICWC
ASLO	97	98.7	100	100	95	100	100	100
NILJ	93	90.9	98.8	86.8	100	89.4	81.25	90.9
NOHI	88	100	100	100	100	100	100	100
KYCO	86	98.7	98.8	97.4	100	100	100	95.5
SAGO	84	100	100	100	100	100	100	100
GIRU	83	85.7	96.3	76.3	95	94.7	56.25	90.9
JAHA	83	84.4	96.3	76.3	95	89.5	68.75	81.8
GRHA	78	85.7	96.3	89.5	75	89.5	93.75	86.4
MASUL	77	88.3	93.9	81.6	90	100	56.25	100
TABE	77	96.1	100	97.4	100	89.5	100	95.5
NABU	76	88.3	97.6	81.6	95	94.7	62.5	95.5
JUCA	75	98.7	97.6	100	100	94.7	100	100
CHNO	73	83.1	95.1	84.2	80	84.2	68.75	95.5
CUBR	73	84.4	90.2	84.2	80	89.5	87.5	86.4
MIKI	72	83.1	93.9	76.3	80	100	50	95.5
JAST	71	81.8	93.9	71	90	94.7	37.5	95.5
MICR	70	89.6	98.8	81.6	95	100	68.75	90.9
SAMA	69	88.3	96.3	78.9	95	100	56.25	95.5
JEHA	68	84.4	96.3	73.7	90	100	43.75	95.5
JABA	64	89.6	92.7	86.8	95	89.5	81.25	90.9
ANHE	63	87	96.3	78.9	95	94.7	50	95.5
AUMI	58	85.7	93.9	81.6	90	89.5	75	86.4
BEDO	56	85.7	98.8	81.6	90	89.5	68.75	90.9
ISPR	53	92.2	98.8	89.5	95	94.7	87.5	90.9
MIBE	51	75.3	97.6	73.7	75	78.9	50	90.9
BECL	50	75.3	97.6	73.7	65	89.5	50	90.9
JAAL	48	94.8	95.1	89.5	100	100	81.25	95.5

Child	Intel	TTC	V	IC	MC	FC	CC	ICWC
CUJA	45	70.1	85.4	71	80	57.9	50	77.3
LEBR	45	93.5	100	94.7	90	94.7	100	90.9
ADBR	42	89.6	95.1	89.5	90	89.5	81.35	95.5
SYNA	42	92.2	98.8	84.2	100	100	68.75	95.5
KAHA	38	88.3	93.9	84.2	85	100	81.25	86.4
RODA	35	83.1	97.6	76.3	90	89.5	75	77.3
JEHAM	29	98.7	98.8	100	95	100	100	100
MATA	29	90.9	92.7	89.5	85	100	87.5	90.9
MOFE	28	87	93.9	86.8	80	94.7	75	95.5
ELVI	27	85.7	90.2	76.3	90	100	56.25	90.9
NICO	26	74	96.3	63.2	75	94.7	37.5	81.8
ALHA	25	97.4	100	94.7	100	100	93.75	95.5
MABI	24	90.9	95.1	89.5	90	94.7	81.25	95.5
ZANA	19	83.1	98.8	78.9	90	84.2	62.5	90.9
ISRO	8	74	93.9	63.2	80	89.5	31.25	86.4
MAWO	2	37.6	89	36.8	35	42.1	25	45.5
RIKL	0	71.4	90.2	57.9	75	94.7	43.75	68.2
<i>M</i>	54.5	86.5	95.9	82.5	88.4	92.3	71	90.6
<i>SD</i>	25.3	10.8	3.4	12.6	12.1	10.9	21.5	9.7

Table 3

R values between each predictor variable and intelligibility scores and intra-variable correlations for each predictor variable.

	Intelligibility	TTC	V	ICC	MC	FC	CC	IC
Intelligibility								
TTC	.491							
V	.382	.623						
ICC	.484	.941	.589					
MC	.485	.895	.553	.747				
FC	.285	.789	.488	.586	.691			
CC	.364	.795	.505	.918	.575	.395		
IC	.51	.855	.554	.784	.749	.741	.482	

TTC- Total Targets Correct; ICC- Initial Consonants with Clusters; V- Vowels and diphthongs; MC- Medial Consonants; FC- Final Consonants; CC- Consonant Clusters; IC- Initial Consonants without clusters

Squared multiple correlation (R^2) and adjusted R^2 values for the best single predictor variable and two- and three- variable models for predicting intelligibility scores.

Table 4

Predictor Variables	R^2	Adjusted R^2
Single Variable: TTC	0.24	0.22
Two-variable Model: IC and CC	0.28	0.24
Three-variable Model: TTC, FC, and CC	0.31	0.25

Appendix A

Beginners Intelligibility Test (Osberger et al, 1994).

List 1	List 2
1 The baby falls.	1 Daddy runs.
2 Mommy walks.	2 The baby cries.
3 The duck swims.	3 The dog eats.
4 The boy sits.	4 The girl drinks.
5 Grandma sleeps.	5 The clown falls.
6 That is a little bed.	6 That is a big bed.
7 The boy walked to the table.	7 The boy walked to the chair.
8 My car is blue.	8 My van is green.
9 He is brushing his teeth.	9 They are playing the drums.
10 She is taking a bath.	10 She is talking on the phone.
List 3	List 4
1 Daddy walks.	1 The bear sleeps.
2 The bunny drinks.	2 Mommy sits.
3 The dog sleeps.	3 The rabbit hops.
4 The girl jumps.	4 The cowboy jumps.
5 Mommy reads.	5 Grandma falls.
6 That is a brown chair.	6 That is a black hat.
7 The boy is on the table.	7 The boy is under the table.
8 My airplane is big.	8 My airplane is small.
9 He is tying his shoe.	9 He is painting the chair.
10 She is brushing her hair.	10 She is cooking dinner.

Appendix B

Monsen – Indiana University Sentences (Osberger, et al., 1994)

List 1	List 2	List 3
1 This house is white.	1 Our car is safe.	1 His boat is white
2 My dog is mean.	2 That lake is deep.	2 My bike is new
3 Can he make any?	3 Can you tell us?	3 Did we call them?
4 Did you find some?	4 Do you want any?	4 Did you buy it?
5 You got a nice haircut.	5 They saw a long sunset.	5 She ate a good hotdog.
6 We made a nice birdhouse.	6 She saw the poor cowboy.	6 We bought a new baseball
7 Can he stop them?	7 Can you start it?	7 Did you steal it?
8 Did she bring it?	8 Have they reached it?	8 Did you try it?
9 My grandmother is beautiful.	9 My television is broken.	9 Her sweater is purple.
10 That elephant was dangerous.	10 That newspaper was interesting.	10 The butterfly is sleeping.