

Measuring the intelligibility of conversational speech in children

PETER FLIPSEN JR

Department of Audiology and Speech Pathology, University of Tennessee, TN, USA

(Received 7 November 2003; accepted 9 November 2004)

Abstract

Conversational speech is the most socially-valid context for evaluating speech intelligibility, but it is not routinely examined. This may be because it is difficult to reliably count the number of words in the unintelligible portions of the sample. In this study four different approaches to dealing with this problem are examined. Each is based on the assumption that it is possible to perceive syllables in unintelligible strings even when the target words are unknown; these unintelligible syllables can then be used to estimate the number of unintelligible words in these samples using at least four different approaches. Preliminary data are presented for each of the four approaches based on conversational speech from two convenience samples including 320 children with normal (or normalized) speech and 202 children with speech delay. Differences among the four approaches are discussed.

Keywords: *Child speech, conversation, intelligibility*

An underlying goal of all human speech communication is to present a message that is intelligible to (i.e., understood by) another human being. Achieving that goal is no simple matter however, because a large number of variables are involved. First, the speaker and listener must be using the same language. Second, the speaker must produce all of the units of that language reasonably accurately, with sufficient amplitude and with appropriate prosodic variation. Third, the listener must have sufficient hearing to decipher spoken language. Fourth, the noise level surrounding the interaction must be low enough to prevent interference with the signal. Making oneself understood in conversational speech is even more involved, because understanding what is being said is not a single event but rather a series of events. Speaking rates and intensities of output vary considerably from utterance to utterance as does accuracy of production, and length and complexity of the utterances. In addition temporary distractions (auditory or visual) place an added burden on the listener. Of course, conversational speech includes elements that actually assist the listener including the linguistic context of the message (creating varying degrees of predictability within each utterance), nonverbal signals (gestures, postures and facial expressions), and the conversational milieu (i.e., topic being discussed, preceding utterances, etc.). The interest level of the listener plays a role as well. Finally, being

personally familiar with the speaker can offer a significant advantage in understanding what is being said even without formal training (Flipsen, 1995). Given this degree of complexity, it should not be surprising that we have yet to develop computers that understand spoken language at a level comparable to human listeners.

Despite the challenges reviewed above, the problem of measuring speech intelligibility warrants study because reduced intelligibility is a central concern with many speech-impaired populations (both children and adults). This concern is manifest in the wide array of studies that have examined speech intelligibility and in the array of procedures available for assessing intelligibility particularly in the speech of children (Kent, Miolo, & Bloedel, 1994). As reviewed by Kent and colleagues, procedures for measuring intelligibility of speech vary on a number of different dimensions, with a primary methodological variable being the linguistic level of the material being evaluated (e.g., single words, sentences, conversation). The current study sought to add to our knowledge base about one of those levels, namely conversational speech.

Quantifying conversational intelligibility

Schiavetti (1992) has noted that there are two kinds of tasks that can be used to quantify intelligibility and both are applicable to conversational speech. These tasks include scaling procedures where the listener estimates the proportion of the intended targets that were understood and word identification tests in which the listener attempts to determine (by transcription in the case of conversational speech) exactly what was said. Scaling procedures are probably the more frequently used of the two in large part because of their simplicity and efficiency. However, Schiavetti presented a number of arguments from the available literature against using scaling approaches for measuring intelligibility. For example, he noted that when using what appear to be interval scales for estimating intelligibility, studies have shown that listeners do not divide the scale equally. Thus, the scale becomes ordinal rather than interval; comparing individuals of differing intelligibility levels or monitoring change across time in a single individual would be problematic with such scales. Schiavetti also noted that confidence intervals around ratings from intelligibility scales (especially in the middle part of the scale) tend to be quite large rendering individual values very difficult to interpret. Word identification tasks suffer from neither of these problems and offer the additional advantage of presenting a starting point for identifying possible sources of intelligibility deficits. Thus, they have the potential to go beyond merely indexing severity (Schiavetti, 1992).

Despite their advantages word identification tasks are not routinely used for analysis of conversational speech. This avoidance likely reflects the fact that quantifying intelligibility in conversation continues to pose a number of procedural problems. For example, there are questions about how to evoke the sample, how long a sample is needed and who should score the sample. Perhaps the most significant problem is that the total number of words being attempted must be known in order to objectively quantify the percentage understood (Kwiatkowski & Shriberg, 1992). Such counts are not difficult when the majority of the speech being produced is understood. For example, if a child says "I want the X one" (X representing something not understood), the listener may not be able to guess what the missing word was but can infer that the unintelligible portion likely only contained a single word. Long strings of unintelligible speech are however much more problematic. If a child produces entire phrases or

clauses in which few or none of the words are identifiable, it is very difficult to reliably count the number of missing words (Shriberg, 1986). It has even been suggested that calculating the percentage of words understood may be impossible in such cases (Connelly, 1986).

The problem of unintelligible speech in children

Unlike written language, spoken language does not include pauses between most words. Conversational speech in particular consists of a largely continuous stream of acoustic information. Listeners parse the speech stream into word units in their heads by identifying the boundaries between words from the available acoustic information, their knowledge of the language, and the context of the conversation. As noted by Cole and Jakimik (1980), “words, then, are the product of speech perception. They exist in the mind of a perceiver, and not in the physical stimulus” (p. 133).

The process of word parsing works quite elegantly when listening to competent adult speakers of one’s native language. Mis-communication occurs only rarely. However, listening to the speech of highly unintelligible children is a very different situation. It is not unlike listening to speakers of an unknown foreign language (Goldstein, 1983) in which knowledge of the lexicon and rules of the language is lacking. In fact, beginning students in speech pathology have been known to describe children with severely delayed speech as “speaking Martian”. Such reactions should not be surprising. Children with delayed speech (as well as very young typically-developing children) lack either the phonemic knowledge to select and assemble the appropriate phones or the phonetic skill to produce them accurately or both. The problem this creates for the listener is that there is insufficient or incorrect acoustic information available in the output to begin the process of identifying word boundaries. If the acoustic information is lacking for too many words, the linguistic context that normally supplements word identification (Cole & Jakimik, 1980; Norris, McQueen, Cutler, & Butterfield, 1997) is also compromised. The listener is then left facing a long, indecipherable acoustic string just as they would be when listening to a speaker of an unknown foreign language. Identifying or even simply counting the number of words present becomes a very difficult task.

Possible solutions to the problem

Lacking the ability to reliably count words in unintelligible speech need not mean abandoning the enterprise. Shriberg and Kwiatkowski (1980) have suggested that even without knowing the words, transcribers can reliably count and record the number of syllables that were produced. Syllable nuclei (i.e., vowels, diphthongs and syllabic consonants) stand out and can be identified from the acoustic signal as peaks of sonority or relative loudness irrespective of the intended message. Shriberg (1986) conducted a review of published data and examined structural statistics from a large number of child speech transcripts and concluded that young children’s speech consists of approximately 75% monosyllabic words and 25% polysyllabic words. Shriberg suggested that each utterance containing unintelligible portions should be examined and in cases where the transcriber is confident about word boundaries (i.e., where prosodic and contextual information provides helpful cues) the unintelligible syllables can be grouped into the words being perceived. Where there is uncertainty about word status the transcriber is advised to group these unintelligible syllables into mono- and disyllabic words based on a

3:1 approximation rule. Thus, in a string of five unintelligible syllables the first three would be transcribed as monosyllabic words and the next two as a disyllabic word. Having demarcated the unintelligible words in all the utterances in the sample, it is possible then to calculate the percentage of words in the entire sample that the transcriber could reliably understand (a measure termed Intelligibility Index [II]; Shriberg, Austin, Lewis, McSweeney, & Wilson, 1997a).

An alternative to the utterance-by-utterance approach used by Shriberg and colleagues (Shriberg & Kwiatkowski, 1980; Shriberg, 1986; Shriberg et al., 1997a) involves approaching the sample as a whole. Rather than grouping the syllables into word units at all, the transcriber would simply indicate the number of unintelligible syllables present. Having done this throughout the entire sample the total number of unintelligible syllables could then be counted. The resulting total could then be divided by the approximate number of syllables per word (SPW) in children's speech to yield an estimate of the number of unintelligible words. The number of unintelligible words in the sample would then be combined with the number of intelligible words to calculate an index of intelligibility. The advantage of this approach is that no assumption is necessary about the "word" status of any particular unintelligible syllable or group of such syllables. The disadvantage of course is that per-utterance data are lost.

When considering this alternative approach the question arises as to what the appropriate value for SPW should be. One approach to dealing with this question was suggested by Yaruss (2000) who was trying to solve the problem of converting words per minute values to syllables per minute values in studies of rate in children with fluency disorders. Yaruss used per-speaker values of SPW to develop a conversion factor; the general approach Yaruss used might be extended to intelligibility calculations. In this case, each speaker would serve as his/her own control. By calculating the number of SPW in the intelligible portions of the sample, the obtained value could then be used to estimate the number of words in the unintelligible portions.

Another approach to answering the SPW question would be to look at the existing literature and ask how many syllables per word are normally found in speech. A number of data sources are informative here. The 3:1 approximation rule suggested by Shriberg (1986) was based on a review of the literature and considerable available data and yields a value of 1.25 (assuming that five syllables yield four words). A study by Flipsen (2005) presented empirical data from the same participants in the current study and concluded that in fact there is a significant developmental progression in syllable length in normal child speech from 3 to 8 years of age. A regression formula was presented.

Goal of the current study

The above alternatives suggest there may be at least four different ways to calculate the intelligibility level of a conversational speech sample: (1) II-Original in which a transcriber takes an utterance-by-utterance approach and groups the unintelligible syllables into words during transcription, (2) II-1.25 in which the number of unintelligible words is calculated by assuming that on average words in children's speech include 1.25 words (as per the 3:1 approximation rule from Shriberg, 1986), (3) II-PS in which each speaker serves as their own control; the SPW in the intelligible portion of the sample is used to estimate the number of words in the unintelligible portion, and (4) II-AN which uses the regression equation from Flipsen (2005) to calculate age-normalized values for SPW to estimate the number of unintelligible words. The goal for the current study was to compare these four

different methods with each other using samples of both typical and delayed speech acquisition.

Method

Speakers

Details on the two speaker groups studied in the present report are provided in Flipsen (2005). Briefly the Normal Speech Acquisition (NSA) Group included conversational speech samples from 320 children (170 males; 150 females) age 3;1 to 8;10 (mean=5;6; SD=1;3). The children's samples met the criteria for normal or normalized speech acquisition in the Speech Disorders Classification System (SDCS; Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997b). Thus, they were all either typically-developing (having been part of study control groups) or had previously been speech-delayed; in the latter case their conversational speech performance was in the normal range (as defined by the SDCS) at the time of sampling for the current study. The SD Group included conversational speech samples from 202 children (145 males; 57 females) age 3;0–8;5 (mean=4;9; SD=1;1). The children's samples had all been classified by the SDCS as SD (Speech Delay). The SD Group had either been recruited through a speech-language pathologist or had been identified as meeting the criteria for SD during the course of a study. A description of the two speaker groups divided by age and sex is presented in Table I in Flipsen (2005).

Conversational speech samples

Conversational speech samples were collected from all the children using procedures described in Shriberg (1986; 1993). All samples included at least 96 intelligible words; for the NSA Group the mean was 198.4 intelligible words (SD=40.0 words), and for the SD Group the mean was 251.8 intelligible words (SD=134.3 words). Note that an integral part of these procedures is that where possible the examiner "glosses" or repeats any utterance by the child that is of questionable intelligibility. Although this potentially gives the transcriber an advantage in knowing as many of the target words as possible, the goal of the original studies had been to examine speech sound acquisition and error patterns in speech-delayed children and not intelligibility per se. Such glossing is of course a common clinical practice when assessing young children.

Transcription and analysis

All of the samples were transcribed using the narrow phonetic transcription conventions described in Shriberg and Kent (1982; 1995). The transcripts were formatted so that analysis could be carried out using the previously described SDCS (Shriberg et al., 1997b). Transcription agreement estimates for the samples and transcribers used in the current study are reported elsewhere (Shriberg & Lof, 1991; McSweeny & Shriberg, 1995; Shriberg et al., 2005).

For purposes of the current study a word was defined as a string of characters in the transcript surrounded by spaces. An intelligible syllable was defined as the presence of any vowel, diphthong or syllabic consonant. Unintelligible syllables were indicated by an * in the transcript. All word and syllable counts of the intelligible and unintelligible words as

well as mixed words (words in which some of the word could be glossed but at least one syllable was unintelligible) were derived using an updated version of Programs to Examine Phonetic and Phonological Evaluation Records (PEPPER; Shriberg, Allen, McSweeney, & Wilson, 2001).

Intelligibility Index (II) values for each participant were calculated by dividing the number of fully intelligible words by the total number of words (fully intelligible+unintelligible). In each case mixed words were assumed to be unintelligible. The number of unintelligible words was calculated four different ways: (1) "Original" in which the transcriber takes an utterance-by-utterance approach and groups the unintelligible syllables into words during transcription, (2) "1.25" in which the number of unintelligible words is calculated by assuming that on average words in children's speech include 1.25 words (per the 3:1 approximation rule from, Shriberg, 1986), (3) "PS" in which each speaker serves as his/her own control; the average number of SPW in the intelligible portion of the sample is used to estimate the number of words in the unintelligible portion, and (4) "AN" which assumes that it is possible to calculate age-normalized values for SPW which are then used to estimate the number of unintelligible words (using the regression equation developed by Flipsen, 2005).

Results

Results obtained for the four versions of Intelligibility Index (II) for the two study groups are shown in Table I. As indicated mean values obtained for the SD Group were approximately 9–10 percentage points below the mean values obtained for the NSA Group. Lower values for the SD Group are as would be expected given their classification as speech delayed children (i.e., the presence of a noticeable number speech sound errors would be predicted to noticeably reduce the intelligibility of their speech). Values for the four versions of II broken down by age group are shown in Tables II and III for the NSA and SD Groups respectively. Examination of the mean values in Tables I, II, and III indicated that the highest mean values were obtained for II-1.25 and the lowest values were obtained for II-original. To ascertain if this pattern obtained across individuals, scores on the four versions were ranked for individual speakers. For the NSA Group the mean ranks were 1.6, 2.6, 2.6, and 3.1 for II-1.25, II-AN, II-PS, and II-original respectively. For the SD Group the corresponding mean ranks were 1.3, 2.5, 2.8, and 3.4 respectively. Thus, across both groups of children the highest values were consistently obtained with II-1.25 and the lowest values were consistently obtained with II-original.

The overall values for the four versions of II (Table I) were then compared separately within each study group using six pairwise comparisons (paired t-tests) with a Bonferoni-corrected p-value of .008. For both study groups all of the comparisons were significantly different from each other except II-PS versus II-AN.

The absolute values of the differences between the four versions of II were then calculated for each sample, and mean differences are shown in Table IV. As indicated the mean values for the differences between II-original and the three alternate versions were consistently greater than the differences in any of the other comparisons. This was true for both groups of children. Mean differences for the NSA Group were consistently smaller than for the SD Group. Looking at the actual differences between II-original and the three alternate versions, the data suggest that 95% of the differences (i.e., mean+/-2 standard deviations) were within approximately 2–3% of each other for the NSA Group and within approximately 8–9% of each other for the SD Group.

Table I. Values obtained for the four different versions of Intelligibility Index (II) for the two study groups overall.

Group	II-original ¹		II-1.25 ²		II-PS ³		II-AN ⁴	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
NSA	97.72 (2.51)	88.42–100.00	98.25 (2.04)	89.38–100.00	98.18 (2.13)	89.51–100.00	98.18 (2.13)	88.87–100.00
SD	87.04 (11.75)	40.17–100.00	89.73 (10.14)	50.51–100.00	89.22 (10.54)	48.90–100.00	89.31 (10.49)	49.61–100.00

¹ Intelligibility index based on the transcriber grouping unintelligible syllables into words during transcription. ² Intelligibility index based on assuming that each unintelligible word includes 1.25 syllables. ³ Per speaker intelligibility index based on calculating syllables per word from the intelligible words in the sample. ⁴ Age-normalized intelligibility index based on calculating syllables per word from the regression equation from Flipsen (2005).

Table II. Values for the four different versions of Intelligibility Index (II) by Age for the NSA Group.

Age	II-original		II-1.25		II-PS		II-AN	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
3 years	95.68 (3.12)	88.89–100.00	97.31 (2.34)	91.75–100.00	97.16 (2.45)	91.45–100.00	97.15 (2.48)	91.38–100.00
4 years	96.82 (3.01)	88.42–100.00	97.31 (2.63)	89.38–100.00	97.18 (2.71)	89.51–100.00	97.18 (2.76)	88.87–100.00
5 years	98.05 (2.02)	89.84–100.00	98.50 (1.77)	91.13–100.00	98.43 (1.87)	90.41–100.00	98.44 (1.83)	90.83–100.00
6 years	98.43 (1.74)	91.67–100.00	98.57 (1.69)	92.17–100.00	98.52 (1.75)	92.27–100.00	98.53 (1.73)	91.97–100.00
7 years	99.51 (.78)	97.36–100.00	99.54 (.77)	97.56–100.00	99.52 (.80)	97.45–100.00	99.53 (.78)	97.52–100.00
8 years	99.01 (1.09)	97.07–100.00	99.24 (.94)	97.12–100.00	99.24 (.94)	97.06–100.00	99.23 (.95)	97.10–100.00

Table III. Values for the four different versions of Intelligibility Index (II) by Age for the SD Group.

Age	II-original		II-1.25		II-PS		II-AN	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
3 years	80.80 (11.91)	52.16–98.81	84.44 (10.86)	56.67–99.53	83.65 (11.28)	54.97–99.50	83.69 (11.26)	55.56–99.50
4 years	88.60 (10.59)	56.56–100.00	90.78 (9.53)	58.52–100.00	90.34 (9.85)	57.09–100.00	90.39 (9.85)	57.40–100.00
5 years	89.41 (11.82)	40.17–100.00	91.55 (9.56)	50.50–100.00	91.15 (10.01)	48.90–100.00	91.28 (9.79)	49.61–100.00
6 years	90.15 (10.61)	66.78–100.00	93.95 (7.25)	73.45–100.00	93.62 (7.52)	72.78–100.00	93.79 (7.43)	72.85–100.00
7 years	85.40 (13.26)	61.10–98.22	88.80 (12.00)	66.35–98.93	88.31 (11.97)	66.37–98.79	88.68 (12.12)	66.00–98.91
8 years	95.32 (n/a)	n/a	94.92 (n/a)	n/a	94.55 (n/a)	n/a	94.90 (n/a)	n/a

n/a=not applicable (n=1; see Table I in Flipsen, 2005).

Table IV. Pair-wise comparison of the differences among the four versions of Intelligibility Index (II).¹

Comparison	NSA Group		SD Group	
	Mean difference (StDev)	Difference range	Mean difference (StDev)	Difference range
II-orig. vs. II-1.25	.71 (.92)*	0–5.15	2.92 (2.88)*	0–12.72
II-orig. vs. II-PS	.67 (.87)*	0–4.74	2.62 (2.62)*	0–11.34
II-orig. vs. II-AN	.68 (.88)*	0–4.85	2.63 (2.66)*	0–12.19
II-1.25 vs. II-PS	.09 (.14)*	0–.89	.55 (.59)*	0–3.16
II-1.25 vs. II-AN	.07 (.10)*	0–.52	.41 (.38)*	0–1.61
II-PS vs. II-AN	.05 (.09)	0–.68	.32 (.39)	0–2.18

¹ Cell entries are percentages. * Statistically significant ($p < .008$).

Discussion

Each of the four versions of II used in the current study represents an estimate of the conversational intelligibility of the children in the study. Findings indicated statistically significant differences among the values obtained (i.e., unique values could be obtained for three of the four versions) for both children with normal speech and those with speech delay.

Smaller differences among the measures were obtained for the NSA Group than with the SD Group. This likely reflected ceiling effects on the values for that group (i.e., many scores approached 100% understood). The measure II-original clearly stands out among the four measured being discussed. Values obtained for II-original were consistently lower than those obtained using the other three approaches. By definition this means that unintelligible syllables were being grouped into shorter words with this approach. Thus, the possibility of longer intended words (e.g., three- or four-syllable words such as banana, alligator) is minimized among the unintelligible portions of the sample when using II-original. By extension, the three alternate versions of II (II-1.25, II-PS, II-AN) would then allow the possibility of longer forms to be more fully accounted for by approaching the sample as a whole (rather than on an utterance-by-utterance basis). On the other hand, one might argue that the utterance-by-utterance approach used in II-original may better represent reality, because it takes into account available contextual and prosodic cues. The above-suggested tendency to group unintelligible syllables into shorter words when using II-original would then lead to the conclusion that these speakers were actually attempting shorter words than might be expected (i.e., shorter than what would be predicted from the regression equation in Flipsen, 2005). The reciprocal of this is that these speakers were avoiding longer words. The fact that this pattern was observed in children in both study groups would suggest that such word avoidance is not limited to children with speech delay. It would also suggest that such avoidance did not always prove advantageous as the words continued to be unintelligible.

One major advantage to using II-original would appear to be practical efficiency. Unintelligible syllables are grouped into words during orthographic transcription (i.e., glossing), and the only counts required are the number of intelligible and unintelligible words. All three of the alternative approaches require counts of the number of unintelligible syllables followed by the extra step of estimating the number of unintelligible words. Calculation of II-PS requires a further step, namely a count of the total number of syllables in the intelligible words. To carry this out reliably would require at least a broad phonetic transcription of the entire sample to identify the vowels,

diphthongs and syllabic consonants. Thus, II-PS would appear to be the least efficient of the measures discussed.

Aside from its lesser practical efficiency, II-PS is also problematic because it assumes that the number of syllables per word in the unintelligible words will always be comparable to the number of syllables in the intelligible words. As discussed in Flipsen (2005) polysyllabic words may be more likely to be unintelligible. In such cases II-PS may overestimate the number of unintelligible words in the sample.

The magnitudes of the differences among the values obtained for the three alternative approaches (II-1.25, II-PS, II-AN) were notably small, rarely more than three percentage points (see Table IV). It raises the question of how such differences compare to measurement error. Standard error of measurement (SEM) data were unfortunately not available for the current data sets, because calculating reliability of transcription involved repeated transcription using a common gloss. SEM data are not commonly reported in studies of intelligibility, particularly those involving conversational speech. Flipsen (1999) reported reliability data for II-original from the conversational speech of six adults with Down syndrome, although no SEM data were derived. Post-hoc calculations suggested an SEM of 1.6% indicating that 68% of the time values for II-original will fall within $\pm 1.6\%$ of the value obtained. This also suggests that 95% of the time values for II-original will fall within $\pm 3.2\%$ (i.e., twice the SEM). Thus, it would appear that none of the three alternative approaches offers any particular advantage for measuring the intelligibility of speech in children.

The particular application being considered may provide a basis for choosing which of these approaches to use. The production of intelligible speech requires more than simply the accurate production of speech segments, which means that the utterance-by-utterance approach of II-original, which allows for the consideration of factors such as prosodic variation and context, might be the most desirable approach overall. However, when working with speaker populations in which prosodic variation is more limited and thus less helpful (e.g., adults with advanced Parkinson's disease, children with autism), or in cases of very low intelligibility where contextual cues may be of limited assistance, one of the alternative approaches may be preferable. As noted above, differences among those three measures were all within the realm of measurement error; as such they would all appear to provide functionally equivalent values. The above noted concerns about II-PS make it the least preferable of the three approaches. This leaves a choice of using either II-1.25 or II-AN which may hinge on whether one is willing to accept a fixed value for SPW or not.

Ultimately any definitive recommendations for choosing among the four approaches must be followed with caution because of the non-random nature of all of the samples used. Ideally a large psychometrically-sound normative sample should be obtained and a more valid regression equation for SPW developed to allow clinicians and researchers to generate more reliable values for both II-PS and II-AN. A more complete evaluation of the different versions of II proposed herein could then be made. The development of such a database would also allow for the development of acceptable norms for intelligibility in conversation against which clinical samples can be compared. Currently the only reliable data on the conversational intelligibility of children that appear to exist reflect parent report estimates (Coplan & Gleason, 1988), and these have yet to be validated against more objective measures.

Findings of the current study suggest that intelligibility of the conversational speech of children can be reliably quantified; clinicians and researchers need not limit themselves to less reliable and less valid scaling methods. Measures such as II (with the four versions discussed herein) represent potential solutions.

Acknowledgments

Many thanks to Jane McSweeney, Lawrence Shriberg, and David Wilson for their assistance with compiling the data set for this project. Thanks also to two anonymous reviewers for some very insightful comments on a previous version of this paper. Preparation of this paper was supported in part by the National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Grants DC00496 (Lawrence Shriberg, PI) and DC00528 (Barbara Lewis, PI).

References

- Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133–163). Hillsdale, NJ: Lawrence Erlbaum.
- Connolly, J. H. (1986). Intelligibility: A linguistic view. *British Journal of Disorders of Communication*, *21*, 371–376.
- Coplan, J., & Gleason, J. R. (1988). Unclear speech: Recognition and significance of unintelligible speech in preschool children. *Pediatrics*, *82*, 447–452.
- Flipsen, P. Jr. (1995). Speaker–listener familiarity: Parents as judges of delayed speech intelligibility. *Journal of Communication Disorders*, *28*, 3–19.
- Flipsen, P. Jr. (1999). Prepausal accuracy in the speech of adults with Down syndrome. *Clinical Linguistics and Phonetics*, *13*, 55–65.
- Flipsen, P. Jr. (2005). Syllables per word in typical and delayed speech acquisition. *Clinical Linguistics and Phonetics*.
- Goldstein, H. (1983). Word recognition in a foreign language: A study of speech perception. *Journal of Psycholinguistic Research*, *12*, 417–427.
- Kent, R. D., Miolo, G., & Bloedel, S. (1994). The intelligibility of children’s speech: A review of evaluation procedures. *American Journal of Speech-Language Pathology*, *3*, 81–95.
- Kwiatkowski, J., & Shriberg, L. D. (1992). Intelligibility assessment in developmental phonological disorders: Accuracy of caregiver gloss. *Journal of Speech and Hearing Research*, *35*, 1095–1104.
- McSweeney, J. L., & Shriberg, L. D. (1995). *Segmental and suprasegmental transcription reliability (Tech. Rep. No. 2)*, Phonology Project, Waisman Center on Mental Retardation and Human Development, University of Wisconsin–Madison.
- Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, *34*, 191–243.
- Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In R. D. Kent (Ed.), *Intelligibility in speech disorders* (pp. 11–34). Philadelphia, PA: John Benjamins.
- Shriberg, L. D. (1986). *PEPPER: programs to examine phonetic and phonologic evaluation records*. Madison, WI: University of Wisconsin–Madison.
- Shriberg, L. D., Allen, C. T., McSweeney, J. L., & Wilson, D. L. (2001). *PEPPER: Programs to examine phonetic and phonologic evaluation records [Computer Program]*. Madison, WI: University of Wisconsin.
- Shriberg, L. D., Austin, D., Lewis, B. A., McSweeney, J. L., & Wilson, D. L. (1997a). The percentage of consonants correct (PCC) metric: extensions and reliability data. *Journal of Speech, Language, and Hearing Research*, *40*, 708–722.
- Shriberg, L. D., Austin, D., Lewis, B. A., McSweeney, J. L., & Wilson, D. L. (1997b). The Speech Disorders Classification System (SDCS): Extensions and Lifespan Reference Data. *Journal of Speech, Language, and Hearing Research*, *40*, 723–740.
- Shriberg, L. D., & Kent, R. D. (1982). *Clinical phonetics*. New York: Macmillan.
- Shriberg, L. D., & Kent, R. D. (1995). *Clinical phonetics*, second edition. Boston, MA: Allyn & Bacon.
- Shriberg, L. D., & Kwiatkowski, J. (1980). *Natural process analysis (NPA). A procedure for phonological analysis of continuous speech samples*. New York: Macmillan.
- Shriberg, L. D., & Lof, G. L. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics and Phonetics*, *5*, 225–279.
- Shriberg, L. D., McSweeney, J. L., Anderson, B. E., Campbell, T. F., Chial, M. R., Green, J. R., et al. (2005). Transitioning from analog to digital audio recording in childhood speech sound disorders. *Clinical Linguistics and Phonetics*, *19*, 335–359.
- Yaruss, J. S. (2000). Converting between word and syllable counts in children’s conversational speech samples. *Journal of Fluency Disorders*, *25*, 305–316.