Intelligibility of conversational speech produced by children with cochlear implants

Peter Flipsen Jr.*, Lana G. Colvard

Department of Audiology and Speech Pathology, University of Tennessee,
425 South Stadium Hall, Knoxville, TN 37996, United States

Received 8 April 2005; received in revised form 13 August 2005; accepted 8 November 2005

Abstract

The intelligibility of conversational speech produced by six children fitted with cochlear implants before age 3 years was measured longitudinally. Samples were obtained every 3 months during periods of 12–21 months. Intelligibility was measured using both an utterance-by-utterance approach and an approach to the sample as a whole. Statistically significant differences were observed between the two approaches, but the differences were all within the realm of measurement error. Findings indicated that intelligible speech emerges quite rapidly in these children. Conversational intelligibility appears to be superior to that reported in the literature for similar children who use hearing aids but not necessarily as good as in children with normal hearing. Both intelligibility measures were significantly correlated with chronological age, hearing age, and amount of implant use, but were most strongly correlated with chronological age.

Learning outcomes: The reader will be able to (1) describe some of the issues involved in measuring speech intelligibility in children with cochlear implants and (2) describe the pattern of outcomes for the intelligibility of speech produced by children receiving cochlear implants before age 3 years.

© 2005 Elsevier Inc. All rights reserved.
1. Introduction

The ability to make oneself understood is critical to most human interaction, and as such, the failure to develop fully intelligible speech may result in a significant handicap. Such failures are unfortunately a common occurrence in individuals with significant hearing loss. Thus, it should not be surprising that speech intelligibility in individuals with hearing impairments has long been an object of study (see Osberger, 1992 for a review).

Investigations of children who have received cochlear implants suggest that intelligibility scores tend to improve as a function of the length of implant use (e.g., Allen, Nikolopoulos, & Donoghue, 1998; Calmels et al., 2004; Chin, Finnegan, & Chung, 2001; Chin, Tsai, & Gao, 2003; Dawson et al., 1995; Inscoe, 1999; Miyamoto, Iler Kirk, Robbins, Todd, & Riley, 1996; Miyamoto et al., 1997; Mondain et al., 1997; Osberger, Robbins, Todd, Riley, & Miyamoto, 1994; Vieu et al., 1998). Results have also shown that intelligibility of speech tends to be superior in those who receive their implants earlier rather than later (Löhle et al., 1999; Loundon, Busquet, Roger, Moatti, & Garabedian, 2000; Miyamoto, Iler Kirk, Svirsky, & Sehgal, 1999; Tye-Murray, Spencer, & Woodsworth, 1995).

2. Conversational intelligibility

Studies in this area have varied in terms of the linguistic level of the material used to examine intelligibility. For example, speakers in these studies have variously produced single words (Chin et al., 2001; Löhle et al., 1999; Mondain et al., 1997; Vieu et al., 1998), sentences (Chin et al., 2001, 2003; Dawson et al., 1995; Miyamoto et al., 1996, 1997, 1999; Osberger et al., 1994; Tye-Murray et al., 1995), narratives (Tye-Murray et al.), or conversational speech (Allen et al., 1998; Calmels et al., 2004; Inscoe, 1999; Löhle et al.; Loundon et al., 2000). Although it has been argued that no one single measure may ever capture intelligibility completely (Kent, 1993), conversational speech is the level of communication that is most commonly used.

At least five previous studies have examined conversational intelligibility in children with cochlear implants. Loundon et al. (2000) presented data from 40 children measured after 2 years of implant use. Using a 5-point rating scale, and considering those who received their implants between age 4 and 12 years, 5% were rated as completely intelligible. Of those receiving their implants before age 4 years, 6% were rated as completely intelligible. Allen et al. (1998) reported findings from 128 children who had all received their implants by age 7 years. The children were rated using the 5-point Speech Intelligibility Rating (SIR) scale. After 3 years of implant use only 3% of their participants were rated as producing fully intelligible speech, but after 5 years 45% were rated as fully intelligible. Calmels et al. (2004) also used the SIR and studied 63 children who received their implants at ages up to 10 years. After 3 and 5 years of implant use 25% and 34%, respectively, were rated as fully intelligible. Using a modification of the SIR (different descriptors were applied to some of the points on the scale), Inscoe (1999) reported that after 3 and 5 years of implant use 7% and 33%, respectively, were fully
intelligible; age of implantation was not specified in that study. Overall, findings from studies that have used rating scales seem to indicate that for children implanted at least before age 10 years, intelligible conversational speech emerges slowly, but after using their implants for at least 5 years 33–45% are fully intelligible. Löhle et al. (1999) took a different approach and attempted to more directly quantify the percentage of speech produced by these children that could be understood (although it was not entirely clear how they specifically arrived at their estimates). For those children who were implanted between ages 9 and 14 years (following 5 years of implant use), all were rated as less than 40% intelligible. Those implanted between ages 5 and 8 years were rated as 30–90% intelligible, while those implanted between ages 2 and 4 years were rated as 60–90% intelligible (the latter two groups used their implants for 24–36 months).

An unresolved issue with these studies is the question of the best approach to quantifying intelligibility (i.e., rating scales or some type of numerical estimate of words understood). Schiavetti (1992) has argued that rating scales are the less desirable of the two approaches. He cites a number of studies in which listeners appear to be unable to equally divide up interval scales for intelligibility. This effectively renders such scales ordinal rather than interval in nature. This means that comparing speakers with each other or with themselves over time is quite problematic as the degree of change being witnessed would not necessarily be uniform across different parts of the scale. Another problem cited by Schiavetti is that confidence intervals around intelligibility ratings can be quite large, particularly in the middle of the scale range. This raises the question of rating scale reliability. Tasks that involve generating numerical estimates (what Schiavetti calls ‘word identification tasks’) appear to suffer from neither of the above problems.

Word identification tasks from conversation (involving transcription to at least the orthographic level) are not without their own set of problems. In particular, Flipsen (in press-b) pointed out that calculating the percentage of words understood requires precise counts of the number of words being attempted. However, such counts may be very difficult to obtain, particularly in long stretches of unintelligible speech. It is not unlike the problem of listening to someone speak an unknown foreign language. It is difficult to identify where one word ends and the next one begins. Flipsen (in press-b) noted that it is possible to use counts of the number of syllables heard (which stand out as peaks of sonority or relative loudness) to estimate the number of words attempted (see also Shriberg & Kwiatkowski, 1980).

3. Age effects?

Regardless of the sampling mode or method of quantifying intelligibility, it is still not clear exactly how much of an advantage it might be to receive an implant early. Two of the above studies provide data that allow for a direct comparison. Findings from Loundon et al. (2000), who examined outcomes after 2 years, do not reveal much advantage, while the study by Löhle et al. (1999) does at least suggest the possibility of an advantage. Both studies included children implanted early, but neither followed those individuals for more than 3 years. This may be crucial as results of the other studies cited above suggest that significant effects may emerge only after 5 years of implant use.
Asking this question raises a related question: what is the most important independent variable to be considered relative to age? Historically hearing age (HA) has been used to evaluate outcomes of intervention for children with hearing impairments. This is the length of time that the child has been receiving amplification. With cochlear implants it is not clear what might be meant by HA. Virtually all children who receive cochlear implants are fitted with hearing aids for some length of time prior to implantation. With the advent of Universal Newborn Hearing Screening (UNHS), identification often occurs very early and hearing aids are fitted soon after identification. HA and chronological age (CA) would then differ very little. A common criterion for implantation is that the potential recipient should have received only “minimal benefit” from a trial with hearing aids. Thus, it could be argued that HA in the traditional sense may be of limited value as an independent variable in this type of research. It is not surprising then that most studies of outcomes in cochlear implantation have focused on amount of implant use (what might be termed ‘post-implantation age’ or PIA). The one exception to this trend is the study by Chin et al. (2003) who reported on a group of 51 children implanted by about age 5 years. Intelligibility was judged using a sentence repetition task, and findings indicated significant correlations between intelligibility and both CA ($r = .71$), and PIA ($r = .65$).

Cochlear implants are now routinely provided to children with hearing impairments as young as 12 months. As age of implantation gets younger, the gap between PIA and CA gets smaller and at a certain point outcomes may not differ appreciably whether one considers PIA or CA. Comparing outcomes relative to PIA with outcomes relative to CA may in fact be a useful way to both examine the effects of age of implantation and to identify optimum implantation age. And because we can never be certain about the “minimal benefit” of hearing aid use during the earliest periods of development, examination of outcomes relative to HA may also be of some value. Such comparisons (with CA, HA, and PIA) may also allow us to determine the optimal basis for developing expectations for the development of intelligible speech in these children.

4. Focus of the current study

The current study involved an examination of conversational speech intelligibility in a group of six children fitted with cochlear implants by age 3 years. This study was unique in that it appears to be the first to use transcription to quantify conversational intelligibility in this population (with the possible exception of the study by Löhle et al., 1999, in which the analysis method was unclear). It also allowed for an examination of the validity of different approaches to quantifying transcription intelligibility in this population. The current study also examined intelligibility performance relative to CA, HA, and PIA.

5. Method

5.1. Participants

Children were recruited for the current study from Child Hearing Services, a clinical services program at the University of Tennessee, Knoxville. The children were selected
based on the following criteria: (a) prelingually deaf (defined as onset of hearing loss before age 3 years), (b) fitted with a multi-channel cochlear implant by age 3 years, (c) use of the cochlear implant for at least 18 months at the onset of testing. This criterion increased the likelihood that the children would be using real speech, rather than babbling, (d) use of spoken language only as their primary communication mode, and (e) receptive vocabulary performance as measured by the Peabody Picture Vocabulary Test-Third Edition (PPVT-III, Dunn & Dunn, 1997) within two standard deviations of their age group mean (i.e., a standard score of at least 70). This latter criterion was used to ensure that the children would have sufficient language comprehension skill to engage in meaningful conversation.

Six children (five girls and one boy) satisfied the selection criteria and participated in the study. As indicated in Table 1, mean age of identification was 8 months (range birth to 15 months). None of these children was identified under Universal Newborn Hearing Screening as such a program does not exist in the state of Tennessee. In addition, none of these children was identified via any “at risk” criteria. Mean age of implantation was 2 years, 4 months (range = 1;8–3;0). Mean age at the start of the study was 5 years (range = 3;9–6;2). Mean length of implant experience at the beginning of the study was 2 years, 8 months (range = 1;11–3;6). Mean score of the PPVT-III was 82.3 (range = 72–99). The cause of the hearing loss was unknown for participants 1–5; participant 6 (the only boy in the group) had a diagnosis of partial agenesis of the cochlea. None of the children had any other known physical, cognitive, or emotional disability. Parents of the children were all reported to have normal hearing. All six of the children had received individualized intervention prior to the start of the study and continued to do so during the course of the study. In all cases the intervention was a completely oral-based approach (i.e., no signing was involved). None of the children had any special educational placements (i.e., all were in regular education classrooms).

### Table 1

<table>
<thead>
<tr>
<th>Participant</th>
<th>Gender</th>
<th>Age of IDa</th>
<th>Age at implantationa</th>
<th>Initial CAb</th>
<th>Initial implant experiencea</th>
<th>Implant type</th>
<th>PPVT-III c</th>
<th>Number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>0;8</td>
<td>2;4</td>
<td>5;2</td>
<td>2;11</td>
<td>Clarion</td>
<td>89</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>0;0</td>
<td>2;6</td>
<td>4;5</td>
<td>1;11</td>
<td>Nucleus-24</td>
<td>99</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>1;0</td>
<td>3;0</td>
<td>6;2</td>
<td>3;2</td>
<td>Clarion</td>
<td>72</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>0;3</td>
<td>2;0</td>
<td>5;5</td>
<td>3;6</td>
<td>Nucleus-24</td>
<td>77</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>1;3</td>
<td>2;7</td>
<td>4;10</td>
<td>2;3</td>
<td>Clarion</td>
<td>81</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>0;11</td>
<td>1;8</td>
<td>3;9</td>
<td>2;1</td>
<td>Nucleus-24</td>
<td>76</td>
<td>6</td>
</tr>
<tr>
<td>Mean (S.D.)</td>
<td></td>
<td>0;8 (0;6)</td>
<td>2;4 (0;6)</td>
<td>5;0 (0;10)</td>
<td>2;8 (0;8)</td>
<td></td>
<td>82.3 (10.0)</td>
<td>6.7</td>
</tr>
</tbody>
</table>

* a Expressed in years;months.
* b Chronological age at start of current study (expressed in years;months).
* c Standard score on the Peabody Picture Vocabulary Test-Third Edition (Dunn & Dunn, 1997).
5.2. Study design

The current study used a prospective, longitudinal design with testing occurring every 3 months. This interval was selected so as to maximize the opportunity to observe changes in intelligibility but minimize the interruption of clinical treatment. The original intention of the study had been to follow each participant for up to 24 months. Actual participation ranged from 12 to 21 months. Differences across the participants resulted from several factors including scheduling difficulties, late enrollment for services at the clinical site being used, participant moving to another treatment facility, and limited availability of the graduate student examiners. The number of samples for each participant is shown in Table 1. Participants were paid US$ 10.00 for participation in each recording session as well as an additional US$ 60.00 at the completion of the study.

5.3. Conversational speech samples

The conversational speech samples used in the current study were evoked as part of a larger protocol lasting 60–90 minutes per session. The PPVT-III was administered during the first session. Samples were evoked inside a single-wall sound-treated booth by either the second author or a trained graduate student clinician. For many of the samples, a parent or the clinician who was providing treatment to the child was also present in the booth and participated in the interaction. The samples were recorded on digital audiotape using a Sony TCM-150 microphone held by a puppet which was connected to a Sony PCM-M1 portable digital tape recorder. The sampling rate on the recorder was set at 48 KHz. Conversational speech was evoked using a variety of topics (e.g., favorite movies or cartoons, current activities in therapy) and materials such as activity pictures from the Bracken Concept Development Program (Bracken, 1998). A target sample size of at least 90 different words was selected because samples of this size have been shown to provide a representative sample of English phonemes and canonical forms (Shriberg, 1986). Narratives were avoided because of concerns about the use of non-typical prosody in the narrative register (Shriberg, Kwiatkowski, & Rasmussen, 1990). The first author monitored each test session from outside the sound booth through headphones and kept a running tally of the different intelligible words produced by the children. In one case, the sample only contained 67 words because the sample was terminated after approximately 25 min in order to avoid fatigue effects. All other samples included at least 92 words. Samples were allowed to run to at least 15 min of conversation to allow for the possibility of later analysis of other aspects of language production in these children. The resulting samples varied in length from 67 to 199 different intelligible words (mean = 139.0; S.D. = 26.1) and included between 65 and 216 utterances (mean = 134.3; S.D. = 35.5).

5.4. Sample transcription

The conversational speech samples were transcribed by a trained graduate student clinician who had completed an undergraduate course in phonetics. Narrow phonetic transcription conventions were used throughout using the system of Shriberg and Kent (1995). Supplemental transcription training consisted of having the student practice
transcription of several conversational samples of delayed speech in children that had previously been transcribed by a clinician with over 20 years of experience in clinical phonology. The initial transcripts were used to provide corrective feedback during the training. Following training, transcription accuracy (effectively a form of inter-judge reliability measurement) was assessed by having the student transcribe two additional samples that had been previously transcribed by the experienced clinician. The student transcriptions were made using the original gloss (i.e., regular spelling target) in order to avoid calculation problems associated with disagreements about the original targets. That is to say, if the targets were judged to be different words, it would not be a fair comparison as it would be virtually impossible to have agreement on the phonemes being produced. Based on a total sample of 564 vowels and 759 consonants, point-to-point agreement for broad transcription was 90.4% for vowels and 87.3% for consonants. Agreement for narrow transcription was 87.3% for vowels and 83.5% for consonants. These values are comparable to those obtained in other studies using phonetic transcription (e.g., Shriberg & Lof, 1991).

The longitudinal nature of the current study created the possibility of familiarity effects during the transcription process. To avoid such effects, the samples were transcribed in random order.

5.5. **Intelligibility measurements**

Two measures of intelligibility were used in the current study. The choice of these measures was made based on findings from Flipsen (in press-b), who compared Intelligibility Index (II; Shriberg, 1986) against three variations of it. All four variations of II are based on the idea of counting syllables heard in unintelligible speech. Each variation takes a different approach to grouping those unintelligible syllables into words. As proposed by Shriberg (1986), Intelligibility Index (hereafter II-Original) involves dealing with the sample on an utterance-by-utterance basis. The transcriber uses any available prosodic and contextual cues to define the number of missing words. In situations where such cues are unavailable or unhelpful, syllables are then grouped into words using a default rule of three monosyllables for each disyllable. The other three versions of II involve approaching the sample as a whole. The transcriber first indicates each unintelligible syllable heard on the transcript (without regard to word status). The total number of unintelligible syllables in the transcript is then converted to an estimated number of unintelligible words using one of three possible numerical estimators. The first of these estimators involves generalizing Shriberg’s 3:1 monosyllable:disyllable rule to conclude that young children’s speech averages approximately 1.25 syllables per word (derived from the fact that five unintelligible syllables would yield four unintelligible words). Flipsen (in press-b) referred to intelligibility values derived using this approach as II-1.25. The second numerical approach (termed the “per sample” approach or II-PS) involves using the number of syllables per word in the intelligible portion of the sample to estimate the number of words in the unintelligible portion (each sample thus serves as its own control). The last approach (termed the “age-normalized” approach or II-AN) involves estimating the number of unintelligible words using a regression equation derived by Flipsen (in press-a). The equation was based on a large sample of normal speech in 3–8-year-old children and results in age-normalized values of syllables per word.
Flipsen (in press-b) reported that differences among II-1.25, II-PS, and II-AN were all within the realm of measurement error. Flipsen therefore recommended using II-Original in those cases where prosodic and contextual cues were likely to be helpful, but choosing between II-1.25 and II-AN in all other cases (II-PS was not recommended because it involved additional analysis). The choice between the latter two measures would be based on whether one was comfortable with assuming that the number of syllables per word was a fixed value (II-1.25) or potentially varied with age (II-AN) in the population being considered.

II-Original and II-AN were used in the current study (it was not initially clear whether prosodic information would be helpful with this population). II-AN was chosen over II-1.25 because it did not assume a fixed value for syllables per word. The transcriber used the II-Original approach during transcription. Values for II-Original were generated by an updated version of Programs to Evaluate Phonetic and Phonologic Evaluation Records (PEPPER; Shriberg, Allen, McSweeny, & Wilson, 2001). Total counts of the number of intelligible words (IW) and the number of unintelligible syllables (US) on all 40 samples were made from the phonetic transcripts by the first author. US counts were then converted to unintelligible words (UW) by dividing by the number of syllables per word. Estimates of syllables per word were obtained using the equation shown below (taken from Flipsen (in press-a):

\[
\text{Syllables per word} = 1.12136 + (0.0012277 \times \text{age in months})
\]

Values for II-AN were then generated using the following equation:

\[
\text{II-AN} = \left( \frac{\text{IW}}{\text{IW} + \text{UW}} \right) \times 100
\]

5.6. Data analysis

Two analyses were undertaken for the current study. First, intelligibility values obtained for the 40 samples were compared across II-Original and II-AN using a paired \(t\)-test. This particular test was selected because the measures were made from the same samples and thus were not independent of each other. The second analysis involved correlating each of the intelligibility values with CA, HA, and PIA to look for developmental trends. Given a set of three correlations with each intelligibility measure, a Bonferroni-corrected \(p\)-value of .017 was used to test for statistical significance.

5.7. Reliability

Following completion of transcription, five randomly selected samples were re-transcribed to check intra-judge reliability. The time between the two transcriptions was at least three months. As was the case during training, the re-transcriptions involved starting with the gloss (i.e., the adult target in regular spelling) generated during the first transcription. Based on a total sample of 1908 vowels and 2812 consonants, point-to-point agreement for broad transcription was 94.9% for vowels and 88.9% for consonants.
Agreement for narrow transcription was 93.9% for vowels and 87.7% for consonants. Again, these values are quite consistent with values obtained in previous studies of phonetic transcription (e.g., Shriberg & Lof, 1991).

The use of the original gloss for measuring reliability of phonetic transcription meant that reliability of glossing (i.e., reliability of the intelligibility measures) could not be calculated using the initial set of re-transcriptions. To get around this problem, five additional samples were randomly selected and re-glossed by the student after completion of the initial transcriptions. The time between the re-glossing and the original transcription was again at least three months. The differences between the first and second glossing ranged from 1.4% to 6.4%. Pooled II-Original values ranged from 76.8% to 97.6% (mean = 86.5; S.D. = 7.6), and the (Spearman) correlation between the two glosses was .90. Taken together this suggested a standard error of measurement (SEM) of 2.4%. Thus, 68% of the time the II-Original values would be within 2.4% of each other, and 95% of the time the II-Original values would be within 4.8% of each other (twice the SEM). Glossing reliability values are rarely reported for cases of unintelligible speech, but the value obtained in the current study is similar to findings in Flipsen (1999) which suggested an SEM of 1.6% for transcripts from a group of six adults with Down syndrome (based on post-hoc calculations reported in Flipsen, in press-b). Point-to-point agreement values for the five samples in the current study ranged from 79.8% to 92.0% (mean = 83.2; S.D. = 5.1).

To assess intra-judge reliability of the counts used to generate II-AN, the first author re-counted IW and US from a randomly chosen set of five transcripts. The re-counts were made at least three months after the original counts. Based on a total sample of 667 utterances, the percentage of exact matches on counts for IW and US were 96.0% and 98.5%, respectively. To assess inter-judge reliability of the counts, a second trained graduate assistant re-counted IW and US from a randomly chosen set of five transcripts. Based on a total sample of 723 utterances, the percentage of exact matches on counts for IW and US were 97.4% and 99.6%, respectively.

6. Results

Across the 40 samples, values for II-Original ranged from 65.6% to 96.5% (mean = 85.7; S.D. = 7.9). Values for II-AN ranged from 69.4% to 97.2% (mean = 87.7; S.D. = 7.1). Differences between the two measures ranged from 0% to 4.2% (mean difference = 1.9; S.D. = 1.2). Results of the paired \( t \)-test revealed that the obtained values

<table>
<thead>
<tr>
<th>Measure</th>
<th>For all six participants</th>
<th>Without participant 2 (see text)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CA</td>
<td>HA</td>
</tr>
<tr>
<td>II-Original</td>
<td>.42*</td>
<td>.48*</td>
</tr>
<tr>
<td>II-AN</td>
<td>.43*</td>
<td>.50*</td>
</tr>
</tbody>
</table>

| * CA, chronological age; HA, hearing age; PIA, post-implantation age (amount of implant use). |
| * Statistically significant (\( p < .017 \)). |
were significantly different from each other ($p < .05$). However, all of the differences were less than 4.8% (twice the SEM reported above) suggesting that the differences were all within the realm of measurement error.

Results of the initial correlational analyses are shown in the left-most columns of Table 2. As indicated, the correlations for both intelligibility measures were statistically significant ($p < .017$) for both CA and HA but not for PIA. The significant correlations accounted for between 17% and 25% of the variance in the relationships.

An examination of the intelligibility data suggested possible ceiling effects for participant 2 that might be affecting the results. Despite having the least amount of implant experience at the beginning of the study, she had an initial II-Original value of 96.5% which represented the highest score for any of the participants at any test time (across the study period her scores ranged from 89.8% to 96.5%). This suggested that participant 2 might be an outlier relative to this group of children. It is noteworthy that participant 2 was identified earlier than any of the other participants, and she also had the highest score on the PPVT-III (see Table 1). A post-hoc examination of clinic records revealed additional norm-referenced language testing that had been administered by the treating clinicians (i.e., not the current investigators) to these six children (see Table 3). Although not all of the children were administered the same test, the tests that were used all evaluated more than just

<table>
<thead>
<tr>
<th>Participant</th>
<th>Test</th>
<th>Standard score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Clinical Evaluation of Language Fundamentals-Primary</td>
<td>81</td>
</tr>
<tr>
<td>2</td>
<td>Preschool Language Scale-Third Edition</td>
<td>115</td>
</tr>
<tr>
<td>3</td>
<td>Clinical Evaluation of Language Fundamentals-Primary</td>
<td>70</td>
</tr>
<tr>
<td>4</td>
<td>Preschool Language Scale-Third Edition</td>
<td>71</td>
</tr>
<tr>
<td>5</td>
<td>Clinical Evaluation of Language Fundamentals-Primary</td>
<td>70</td>
</tr>
<tr>
<td>6</td>
<td>Scales of Early Communication in Hearing-Impaired Children</td>
<td>Receptive 67.5</td>
</tr>
</tbody>
</table>

Fig. 1. Chronological age vs. II-AN for 5/6 participants (excluding participant 2).
Looking at Table 3, the performance of participant 2 is striking in how different it was compared to the other five participants (being at least 2 standard deviations higher than any other participant). Because of concerns that the data for participant 2 might have skewed the current findings, the correlations were repeated with the five other participants only (a total of 35 samples). Results of the second set of correlations are shown in the right-most columns of Table 2. As indicated, the correlations with the two intelligibility measures were now significantly associated with all three age measures. The correlations accounted for between 24% and 44% of the variance in the relationships. The strongest obtained relationship (CA versus II-AN for 5/6 participants) is illustrated in Fig. 1.

7. Discussion

Conversational intelligibility for six children with cochlear implants was measured two different ways, the utterance-by-utterance approach of II-Original and the overall sample approach of II-AN. Values for the two measures were statistically significantly different from each other, but all of the differences fell within the range of measurement error. This suggested that the differences observed between the two measures had little practical significance. However, because none of the scores obtained in the current study were below 65%, one cannot rule out the possibility that ceiling effects may have affected the results. Flipsen (in press-b) noted that for children with Speech Delay of Unknown Origin, in which II values reached as low as 40%, some of the differences between II-Original and II-AN exceeded 12% which is considerably larger than measurement error.

Findings obtained in the current study are consistent with estimates obtained by Löhle et al. (1999) who reported speech that was 60–90% intelligible for children implanted between 2 and 4 years of age. Direct comparison of the current findings with studies that used rating scales is problematic because of the stark differences in the dependent variables examined (i.e., percentage words understood versus percentage of individuals with various levels of intelligibility).

The fact that all six of the participants developed conversational speech that was more that 65% intelligible may indicate that cochlear implants may yield much more consistent and better outcomes than have historically been expected for children with severe and profound hearing impairments. For example, Monsen (1978) noted that the speech of children with severe or profound hearing impairments (without cochlear implants) can vary widely from being quite intelligible to being extremely unintelligible even after many years of intervention. Kyle (1977) reported on a group of 15–16-year-olds with hearing impairment (n = 195) and noted that those with hearing losses greater than 85 dB were less than 9% intelligible on average (based on sentence level material rather than conversational speech).

With the exception of participant 6 (who was 3;9 at time 1), all of the children in the current study were chronologically at least 4-year-old throughout the study and thus (if they had normal hearing) would have been expected to be producing fully intelligible speech (Coplan & Gleason, 1988). With the exception of participant 2, who obtained a value of 96.5, values for II-Original obtained at initial testing ranged from 68.4 (participant 6)
to 87.4 (participant 4). Thus, with the possible exception of participant 2, children in the current study were producing less than fully intelligible conversational speech. This finding is consistent with that of Chin et al. (2003) who compared children with normal hearing and those with cochlear implants on a sentence-level intelligibility task. Chin et al. reported significantly less intelligible speech in those with cochlear implants.

Relative to implant experience however, a low value of 68.4 combined with at least 23 months of implant experience suggests that for children who are implanted by age 3 years, intelligible speech begins to emerge quite rapidly in the first 2 years of implant use. This is fairly consistent with the pattern for children with normal hearing for whom intelligibility is expected to reach at least 50% by age 2 years (Coplan & Gleason, 1988). However, normal hearing children are expected to achieve 100% intelligibility by age 4 years, and none of the participants achieved this goal even after over 4 years of implant use. This suggests that the rate of growth in intelligible speech may slow somewhat during the next 2 years of implant use.

It is noteworthy that despite a relatively high level of conversational intelligibility in these children, Flipsen and Kloos (2004) reported that 5/6 of the current study participants (all but participant 2) were unable to accurately imitate production of all of the point vowels (/i, u, æ, ə/) by the end of the study. This suggests that detailed examination of the factors contributing to the development of intelligible speech in this population is warranted.

The presence of the outlier in the current study group raises questions about why some children appear to do so well with cochlear implants and others do not. Participant 2 performed noticeably better than any of the other five participants on both language and speech production testing. The fact that she was identified the earliest may be relevant. All of the participants in the current study were fitted with hearing aids after initial identification, and all but participant 5 were aided binaurally. All participants wore the hearing aids until they received their implants. Because of earlier identification, participant 2 was able to take advantage of auditory input from an earlier age. The other five participants went through periods of from 3 to 15 months immediately after birth when they were experiencing significant auditory deprivation. Another possible explanation for the better performance of participant 2 is that she may have been precocious under any condition (i.e., she was simply genetically predisposed to do very well).

When examining the data for all six participants, the correlational analyses revealed significant associations between the intelligibility of conversational speech and CA and HA, but not with PIA. When the potential outlier was removed from the analysis, all of the correlation coefficients became larger and the correlations with PIA were then also statistically significant. Combined with very high overall levels of intelligibility, this suggested that participant 2 was indeed an outlier. Assuming this is true, like previous investigations, the speech produced by the children with cochlear implants in the current study improved significantly with CA, HA, and PIA. Given that all three age indices were significantly associated with intelligibility outcomes, and that the strongest association was with CA, this suggests that as long as implantation occurs before age 3 years, we may want to focus on chronological age to set expectations for the production of intelligible conversational speech.
Findings of higher correlations between intelligibility and CA compared to PIA are consistent with the findings reported by Chin et al. (2003). Interestingly, the magnitudes of the correlations are quite similar in the two studies.

Generalization of these conclusions to the entire population of children with cochlear implants does need to be made with considerable caution. Only six children participated in the current study and they may not necessarily represent this population adequately. This possibility is illustrated by the receptive vocabulary (i.e., PPVT-III) scores obtained; all were within 2 standard deviations of the mean for normal hearing children. This exclusionary criterion may be too narrow and may have excluded too many children. This possibility is supported by findings from Spencer and Bass-Ringdahl (2004) in which 19 children fitted with cochlear implants at 12–29 months of age achieved standard scores on the PPVT-III ranging from 42 to 102 after 36 months of implant use. Generalization of the findings in the current study should also be tempered by the previously noted possibility of ceiling effects which may have masked some variability in the outcomes.

8. Conclusions

Findings from this study suggest that it is possible to use transcription to document intelligibility of the conversational speech produced by young children fitted with cochlear implants. Intelligible speech appears to emerge quite rapidly in children implanted before age 3 years. Outcomes appear to be superior to historical findings for children with severe or profound hearing impairments who used hearing aids but not necessarily as good as children with normal hearing. At least for children implanted by age 3 years, chronological age may be the best basis for setting expectations for the development of intelligible conversational speech.

Acknowledgments

Many thanks to Julie Beeler, Velvet Buehler, Molly Erickson, Mark Hedrick, Marge Hudson, Gayla Hutsell, Kim Jenkins, Leslie Jolles, Jane McSweeney, Rhonda Parker, Patrick Plyler, Lawrence Shriberg, and Saneta Thurman for their assistance with this study. Special thanks to Michelle LaGesse for assistance with participant testing and to Lisa Moss for transcribing the conversational speech samples. Many thanks also to the children who participated and their parents. A portion of the data in the current study were used as part of a Master’s degree thesis prepared by the second author at the University of Tennessee, Knoxville. Portions of the data used in this study were also presented as part of a poster presented at the 2003 ASHA Convention.

Appendix A

Self-study questions
1. Measuring conversational intelligibility using transcription is preferable to rating scales because:
   a. ratings on rating scales differ greatly from one listener to another.
   b. listeners do not tend to treat all parts of a rating scale equally.
   c. it’s easy to count the number of words not understood.
   d. all of the above.
   e. only a and b.

2. Hearing age (HA) in someone who uses a cochlear implant refers to:
   a. the amount of time they have been using amplification of any kind.
   b. the amount of time they have used the implant.
   c. the amount of time since they were identified as having a hearing loss.
   d. the amount of time between first receiving a hearing aid and receiving a cochlear implant.
   e. none of the above.

3. Counting words in highly unintelligible speech is best done by:
   a. counting the words directly.
   b. counting the number of (unintelligible) syllables attempted.
   c. making an informal estimate after listening to the whole sample.
   d. measuring the speaker’s fundamental frequency.
   e. none of the above.

4. In the current study, intelligibility was found to be correlated with:
   a. hearing age.
   b. chronological age.
   c. post-implantation age.
   d. all of the above.
   e. only b and c.

5. Intelligibility of conversational speech in children who receive cochlear implants before age 3 years:
   a. appears to increase fairly rapidly in the first 2 years after implantation.
   b. appears to reach 100% at the same time as normal-hearing children.
   c. never exceeds 60%.
   d. usually plateaus at 75%.
   e. cannot be reliably measured.

References


