RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 23 (1999)
Indiana University

Speech Intelligibility of Pediatric Hearing Aid Users[1]

Mario A. Svirsky, Steven B. Chin, Matthew D. Caldwell and Richard T. Miyamoto[2]

DeVault Otologic Research Laboratory
Department of Otolaryngology-Head & Neck Surgery
Indiana University School of Medicine
Indianapolis, IN 46202

[1] This study was supported by NIH-NIDCD grants DC00064 and DC00423. Terri Kerr assisted with data management; Allyson Riley, Amy Robbins, and Susan Sehgal collected the data, and Karen I. Kirk and Susan Sehgal gave useful suggestions for the manuscript. All this help is gratefully acknowledged.
[2] All of the authors are also affiliated with the Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, IN 47405.

Speech Intelligibility of Pediatric Hearing Aid Users

Abstract. This study examined the speech intelligibility of profoundly, prelingually or congenitally deaf children who use hearing aids**.** Children were one to fifteen years old and they were classified into subgroups according to residual hearing (PTA between 90 and 100, 100 and 110 or greater than 110 dB HL) and communication mode (either oral or total communication). They read lists of standard sentences which were played back to panels of three "naive" listeners who were not familiar with the speech of the deaf and who did not know which subgroup the children belonged to. The data revealed a strong significant trend toward higher intelligibility for children with more residual hearing, and a significant trend toward higher intelligibility for users of oral communication than for total communication users. However, the latter trend was much more pronounced for some ranges of residual hearing than for others, and it may have been partly due to a sampling effect. A third trend showed significantly higher intelligibility levels at older ages, but this was particularly pronounced for children with PTA's between 90-100 dB HL, and for the majority of oral communication users (and only a few total communication users) with PTA's between 100-110 dB. These results suggest that the amount of residual hearing (possibly in interaction with the communication mode used by the child) may be an important factor in the development of intelligible speech.

Introduction

A connection between prelingual hearing impairment and atypical speech production has been observed for millennia (e.g., Hippocrates, 1853, p. 609), and it is known that the speech of prelingually deafened children can be very difficult to understand. Nevertheless, throughout history, people have attempted, with varying degrees of success, to impart to children with hearing impairment the skills necessary to allow them to communicate successfully in a world so dependent on spoken language. Among the myriad measures used to assess speech production in children (e.g., phonemes correct, suprasegmentals correct, aerodynamic deviance), the one with the highest face validity as regards the need to communicate using spoken language is overall speech intelligibility. Measures of overall speech intelligibility cut to the quick, addressing the important question, "Can this child be understood?"

The question of whether a child with hearing impairment can be understood appears a simple question, but stating the question, and, perhaps more importantly, answering the question, do not by any means lack scientific subtlety. The question "Can this child be understood?" immediately raises obvious follow-up questions: "By whom?" and "How much of the time?" and "At what level?" The scientific literature on speech intelligibility among children with hearing impairment is not uniform in its approaches to these questions, and how best to address the question of speech intelligibility can be as much a matter of practical considerations as of theory.

Measuring the speech intelligibility of children with hearing impairment has generally relied on two types of tasks: (1) rating scales and (2) write-down procedures (Samar & Metz, 1988). In the first type of task, listeners make explicit judgments about the talker's (overall) speech intelligibility by assigning numerical values to samples of speech (e.g., the NTID rating scale: Subtelny, 1977), whereas in the second type of task, listeners "write down what they thought each child said" (Monsen, 1981). As Metz, Schiavetti, & Sitler (1980) point out, write-down procedures appear to have higher face validity than a rating task, because results depend on the listener actually and literally understanding what has been said. Moreover, write-down procedures are relatively insensitive to vocal qualities of speech; these may contaminate responses on a rating scale, but they need not of necessity degrade message intelligibility (Samar & Metz, 1988). On the other hand, write-down protocols are time consuming and labor intensive and therefore expensive (but see Samar & Metz, 1988), whereas rating tasks are viewed as relatively “quick and easy” (Metz et al., 1980). Some considerations making write-down intelligibility assessments difficult to administer are that listeners must respond to every word heard, and that individual listeners should not hear the same word or sentence more than once, to avoid an order effect from increased familiarity with the spoken material. Disadvantages of rating scale tasks are the relatively lower face validity of the protocol, as noted above, and the lack of material for phonemic analysis of listener misidentifications.

A second consideration in speech intelligibility assessment is the type of listener judge, roughly, experienced vs. inexperienced (naive). “Experienced” generally means experienced with the speech of people with hearing-impairments, and such experienced listeners may include educators of the deaf, audiologists, and speech-language pathologists (e.g., McGarr, 1983). Naive listeners are those who have had little or no experience with the speech of the hearing impaired. There is often a connection between the type of task involved in speech intelligibility assessment and the type of listeners involved in the protocol. Rating scales, especially equal-interval scales, are often administered to experienced listeners (e.g., Subtelny, 1977), whereas inexperienced listeners very often participate in write-down tasks (e.g., John & Howarth, 1965).

A third parameter in the assessment of speech intelligibility is the type of speech material used, including both elicitation protocols and the linguistic level of analysis. One parameter in elicitation procedures is that of spontaneous speech vs. read speech (Smith, 1982). If spontaneous speech is elicited, a semantic/pragmatic context is generally established, such as a set of pictures either for description or as basis for a story (on the other hand, context can be an independent variable within a single investigation e.g., Becker, Schildhammer, & Ruoß, 1994). Read speech is usually elicited as a set of full sentences (e.g., Hudgins & Numbers, 1942; Maasen & Povel, 1985; Monsen, 1978, 1983; Osberger & Levitt, 1979), although the SPINE (Monsen, 1981) elicits read versions of isolated words. One variation of read speech is used with preliterate children, who may be asked to repeat sentences spoken to them by an examiner (e.g., Osberger, Robbins, Todd, & Riley, 1994). The level of analysis for measuring speech intelligibility also varies across studies and instruments. Rating scales generally apply to linguistic units at least as large as sentences and ranging up to entire passages (e.g., The Rainbox Passage, from Fairbanks, 1960). The NTID Speech Intelligibility Scale (Subtelny, 1977), for instance, rates connected spontaneous discourse on a 5-point scale ranging from “Speech cannot be understood” (1) to “Speech is completely intelligible” (5). Analysis units for identification tests are generally either words or sentences. Words can be tested either in isolation (e.g., the SPINE test: Monsen, 1981) or in phrasal or sentential contexts (e.g., John & Howarth, 1965; Maassen & Povel, 1985; Markides, 1970; McGarr, 1983). Words in phrases or sentences can be scored on the basis of keywords only (e.g., McGarr, 1983; Smith, 1975; or CID Everyday Sentences [Hirsch et al., 1952]); all words equally (e.g., John & Howarth, 1965); or all words, with weighting (e.g., Monsen, 1978, 1983, with different scores for individual words based roughly on frequency and predictability). When whole sentences or phrases are the units of analysis, listeners must perforce hear each word correctly; in the scoring procedure of Hudgins and Numbers (1942), for example, “no credit was allowed for partially correct auditions” (p. 302).

Aside from construct variables in speech intelligibility assessment, research on speech intelligibility may also differ according to subject characteristics. Some important subject variables in studies of the speech intelligibility of children with hearing impairments are chronological age, degree of hearing loss (conversely, amount of residual hearing), age at onset of hearing loss, duration of hearing loss, and communication mode (oral communication vs. total communication [i.e., spoken and signed]).

In the present study, we assessed the speech intelligibility of hearing aid users in a write-down task (because of its higher face validity) with naive listeners. Write-down procedures date to at least Hudgins and Numbers (1942), an early study using this procedure to examine the speech intelligibility of deaf children. In that study, experienced, rather than naive, listeners served as judges, because the authors believed that inexperienced listeners would be distracted by atypical voice characteristics and would "thus lose much of the content" (p. 301). One-hundred ninety-two children participated and were divided into three groups according to the degree of hearing loss, using the classification of Guilder and Hopkins (1936). The children were between 8 and 20 years of age, and all were being educated at oral schools. Phonographic recordings were made of the children reading 10 sentences (e.g., "Sally likes to swim"). Listeners either were teachers of the deaf or were training to be teachers of the deaf. An average of seven listeners heard each of the sentences. Listeners heard each sentence three times and were instructed to "write down what you think the child says after each reproduction"; listeners were allowed to correct what they had written on previous trials. Responses were scored according to whole-sentence-correctness, that is, regardless of the number of words correctly or incorrectly transcribed. Mean intelligibility across subjects with all degrees of hearing loss was 29.2%.

Whereas Hudgins and Numbers (1942) used experienced listeners and scored whole sentences, John and Howarth (1965) used inexperienced listeners and scored words in a study of the effects of time distortions on the speech intelligibility of deaf children. Twenty-nine children each read a sentence both before and after intensive speech training. Five children had hearing losses below 80 dB, five with losses of 80-89 dB, five with losses of 90-99 dB, and fourteen with losses of 100 dB or more. Twenty "lay" listeners listened to tapes of the 29 sentences both before and after training and were instructed to "write down as much as they could of each phrase" (p. 131). Listeners' responses were scored according to the number of words heard correctly. Across the entire group of 29 children, which included 10 with less than profound hearing losses, the percent words correctly understood by listeners (the measure of intelligibility) was 19% before training. Using methods adapted from John and Howarth (1965), Markides (1970) examined the speech intelligibility of 58 deaf children (mean hearing loss = 95 dB). Stimulus materials were five unrelated pictures for description by the participating children, who were tape-recorded. These recordings were played to panels of listeners consisting of three university students "who were totally naive with regard to speech of deaf children" (p. 128). Listeners were instructed to write down as much as they could of what the child said; scores for each child were means of the number of words correctly transcribed by the three judges. The mean score across children for responses by these naive listeners was 160 words correctly transcribed of 825 words produced, or 19.4%.

Smith (1975) examined the relationship between residual hearing and speech production in deaf children. Forty children participated in the study; all but three had pure tone average (PTA) thresholds of at least 92 dB in the better ear. Each child recorded a list of 20 sentences, and each sentence was later played for three listeners, "without significant previous experience in hearing the speech of deaf persons" (p. 797). Listeners were allowed to hear sentences up to twice each and then wrote down what they thought the children had said. Identification responses were scored word by word; scores for each child were the percentage of keywords correctly understood. Across children and listeners, percent correct scores ranged from 0 to 76.1%, with a mean of 18.7% (SE = 3.2). Monsen (1978) compared intelligibility and acoustic measures in the speech of hearing-impaired adolescents. Thirty-seven adolescents ranging in age from 12;6 (years; months) to 16;6 participated; 27 of the subjects had PTA thresholds better than 95 dB in the better ear. Speech materials were simple sentences containing only common monosyllables and spondees; these were read by the subjects from typewritten copies and audio-recorded. Two successive repetitions of each sentence were played through a loudspeaker to naive listeners "who had never before knowingly heard the speech of a hearing-impaired person" (p. 202). Listeners were instructed "to write down in normal English orthography what he though each subject said, to guess if necessary..." Scoring was based on a maximum value of 10 for each sentence, with scores for individual words assigned according to frequency in the language (low scores = high frequency and predictability; high scores = low frequency and predictability). There were no partial scores for partially correct word responses. Across talkers and listeners, mean intelligibility ranged from 31.3% correct to 99.9% correct, with an average intelligibility score of 76.7%.

Monsen (1981) described "an easy way to use an accurate test for the intelligibility of the speech of severely hearing-impared speakers" (p. 845), called the Speech Intelligibility Evaluation (SPINE). The full protocol as described requires pretesting rehearsal with the examiner, although the test phase itself does not appear to require an experienced listener. During testing, children are shown cards on which are printed one item of a four-item set (e.g., <u>feel</u>, <u>fill</u>, <u>fail</u>, <u>fell</u>). The child is asked to say the word on the card, and the listener must decide which of the four words in the set the child has said; the procedure is thus a four-alternative closed-set task. The entire test consists of 10 such sets; some are true minimal sets (as the one cited above), whereas others are overlapping minimal pairs (e.g., <u>ten</u>, <u>den</u>, <u>ton</u>, <u>done</u>). To validate the SPINE, results were compared with results from a write-down procedure using inexperienced listeners. Testing and scoring of the latter protocol were as in Monsen (1978). The SPINE was administered to 42 hearing-impaired children, including 34 classified as profoundly deaf and 8 as severely deaf. Additionally, the same children produced sentences, which were audiotape-recorded and played for 15 listeners. Scores on the SPINE test, which was admnistered by the author, ranged from 43% correct to 93% correct (mean = 77.9%, median = 79.0%), and scores from the write-down procedure ranged from 27% to 100% (mean = 78.7%, median = 82.3%); correlation (Pearson's product-moment) between the two sets of scores was $r = +.86$.

A study by McGarr (1983) compared the intelligibility of deaf speech to experienced and inexperienced listeners. Twenty profoundly deaf (group mean PTA = 98.6 dB) children, aged 8-10 years and 13-15 years, participated in the study. Test materials were 36 monosyllabic words taken from Smith (1975). The children produced these words both in isolation and embedded in sentences; productions were audio-recorded for subsequent hearing by listeners. In a third listening condition, the words in sentences were excised and presented to listeners as isolated words. Listeners were 60 experienced and 60 inexperienced listeners, the latter defined as persons "with no previous experience in hearing the speech of the deaf" (p. 452). For words in sentential contexts, listeners wrote down the entire sentence, although only the test words were used in scoring. For words produced and heard in sentences, the mean percent correct score for experienced listeners was 41% and for inexperienced listeners 30%. For test words produced and heard in isolation, the mean percent correct score was 29% for experienced listeners and 23% for inexperienced listeners. For words both in sentences and in isolation, differences in scores between experienced and inexperienced listeners were statistically significant.

Monsen (1983) examined a number of variables relevant to the study of speech intelligibility in hearing-impaired talkers: presence or absence of a verbal context, auditory-only or audio-visual presentation, number of presentations, grammatical complexity, and listener experience. Subjects in the study were 10 hearing-impaired adolescents, ages 11;7 (years; months) to 15;3. Eight of the subjects had PTAs worse than 95 dB (mean = 104 dB); PTAs for the remaining two subjects were 83 and 88 dB. Subjects recorded 160 sentences on both audio- and videotape, which were later played to both experienced and inexperienced listeners. Listeners were instructed to "write down as much of each sentence as could be understood" (p. 289), and scoring was based on a total of 100 points for each sentence, individual words being assigned points according to frequency of occurrence in the language (cf. Monsen, 1978). Across all listeners and talkers, percent correct was 79%. Across all talkers, experienced listeners scored 84% correct, whereas inexperienced listeners scored 74% correct; this difference was statistically significant ($p < .05$).

Becker, Schildhammer, and Ruoß (1994) examined speech intelligibility in 23 children (ages 9;6 [years;months] to 10;6) and adolescents (ages 12;6 to 15;4). Eighteen of the subjects had PTA thresholds above 90 dB, and five had PTAs between 85 and 90 dB. Speech materials consisted of descriptions of a picture story, which were audiotape-recorded and later played for two groups of listeners. Experienced listeners were six project personnel, and inexperienced listeners were 48 university students. The recordings of the stories were first transcribed by the experienced listeners and later by the inexperienced listeners. Because the specific speech material produced could vary from talker to talker, the transcriptions from the experienced listeners were considered to be "correct"; transcriptions by the inexperienced listeners were subsequently compared to these to generate outcome measures of intelligibility. Results compared percent correct words as a function of talker age and presence/absence of context. Additionally, results were categorized according to both type of word involved (noun, main verb, function word) and type of response (correct, substitution with same-class word, substitution with different-class word, unintelligible). Results indicated that the naive listeners understood more if there was a context. In general, the adolescents were more intelligible than the children. For example, mean percent correct nouns was 40% for the adolescents and 27% for the children; this difference was statistically significant (F (1, 31) = 4.55, $p < .05$).

Various authors (e.g., Gold, 1980; Osberger, Maso, & Sam, 1993) have cited 20% intelligibility as a consistent finding in the literature on the speech intelligiblity of children with hearing impairment. This estimate is based on, among others, Brannon (1964), John and Howarth (1965), Markides (1970), and Smith (1972, 1975), all of whom used inexperienced listeners. Other studies, also using inexperienced listeners, have reported slightly higher intelligibility (e.g., Becker et al., 1994; Maassen & Povel, 1985; McGarr, 1983). The by-now classic figure of 29% from Hudgins and Numbers (1942) is nevertheless for a study involving experienced listeners. The largest discrepancy in the literature, however, is the one existing between Monsen (1978, 1981, 1983) and just about everyone else. Monsen (1978) notes this discrepancy, citing three studies (John & Howarth, 1965; Markides, 1970; Smith, 1973) as representative of the common result of approximately 20% intelligibility. In the case of Smith (1973), for instance, intelligibility scores ranged from 0% to 76.1%, with a mean of 18.7%. The average in Monsen (1978) across severely and profoundly hearing impaired subjects was 76.7%, that is, more than the maximum score from Smith (1973).

Monsen (1978) further notes "marked differences in the subjects' ages, hearing levels, the recording techniques, materials spoken, listeners, scoring techniques, and so forth" (p. 215). Because of differences from study to study in such basic characteristics as independent and dependent variables, he says that "in one sense, the notion of an average intelligibility figure for such speakers is rather meaningless" (p. 215), yet notes further that the concept of "average intelligibility" can be useful for specific purposes. In attempting to explain the discrepancy between his own work (Monsen, 1978, specifically, but also the later reports, Monsen, 1981, 1983) and that of others, Monsen concentrates on differences in the sentences produced by the subjects. He notes that the sentences in Smith (1973) were more than twice as long on average (10.5 vs. 4.5 syllables) as those used in Monsen (1978), contained more polysyllabic words, and were syntactically more complex. In Markides (1970), there were no standard sentences (children described five pictures), and in John and Howarth (1965), speech materials consisted of children's spontaneous utterances. In the present study, both BIT and Monsen sentences (see below, Methods) were used, which contain on average 4.5 and 5.2 syllables per sentence, respectively.

As indicated above, the communication mode employed by children with hearing impairment is a variable with potentially important effects on the development of intelligible speech. "Oral communication" (OC) educational programs emphasize speech and auditory skill development. All children use hearing aids, and the use of manual signs is not encouraged. "Total communication" (TC) programs use a simultaneous combination of speech and signs. Although much of the manual lexicon is borrowed from American Sign Language (ASL), the manual language used in TC programs encodes the words, morphology, and syntax used orally, which is why it is called “signed English”. Studies of the oral communication skills of children in OC and TC programs have yielded strikingly contradictory results. Extensive literature reviews (Caccamise, Hatfield & Brewer, 1978; Wilbur, 1979) conclude that the use of TC is not detrimental to the development of speech skills. Indeed, Caccamise et al. propose that the use of TC may indeed facilitate the development of speech skills. On the other hand, other studies found a significant advantage in speech intelligibility for children enrolled in OC programs. A comparison of the performance of adolescents with hearing impairment from oral and total communication education settings on the SPINE (Monsen, 1981) was reported in Geers and Moog (1992). This study is of particular interest because the data were analyzed according to two different independent variables: communication mode (i.e., oral or total communication) and degree of hearing loss. The sample consisted of 227 16- and 17-year-olds with PTAs greater than or equal to 80 dB HL, including 100 subjects educated in oral programs and 127 subjects in total communication programs; the latter group included 64 subjects with deaf parents (TC-DP) and 63 with normally hearing parents (TC-HP). Additionally all subjects were divided into four groups according to best binaural PTA threshold: 80-90 dB HL, 91-100 dB HL, 101-110 dB HL, and >110 dB HL. For subjects in oral communication programs, group mean SPINE scores were as follows: 80-90 dB: 93%, 91-100 dB: 87%, 101-110 dB: 83%, >110 dB: 73%. The group mean score for TC-DP students with thresholds in the 80-90 dB range was 69%; for TC-HP students in the same threshold range, the group mean score was 77%. All other TC subjects had mean scores below 60%. Post-hoc comparisons of means showed a significant advantage for students in oral programs over those in total communication programs, and no difference between the two total communication groups.

There are two factors that complicate the interpretation of some of these studies and that may explain some of the discrepancies among them. Firstly, the listeners may be aware of the child’s status and this knowledge may influence the outcome of the study. This is particularly important when the dependent variable is a rating determined by a listener who has a strong belief in a particular hypothesis under study, for example, the superiority of total communication over oral communication training methods, or vice versa. To avoid this problem, in the present study we used naive listeners who were not aware of the speaker’s communication mode or amount of residual hearing. Likewise, the experimenters who conducted the listening sessions were unaware of these details. Second, the choice of total communication or oral communication for a particular child is not an entirely random process. In some parts of the country only one type of program may be available, with the result that geography plays an important role in the selection of communication mode for a deaf child, and becomes a confounding factor. Additionally, children may be steered towards one type of program depending on their aural/oral skills as perceived by their clinicians. The use of TC may be recommended more often in the case of children with poorer aural/oral skills, to ensure that at least some linguistic communication occurs, via the use of sign. In addition, children who start using one type of approach may switch to a different one depending on their success (or lack thereof) in developing aural/oral skills. Children who perform very poorly in OC programs may switch to TC programs and, conversely, children in TC programs who develop very good speech skills may switch to OC. This may introduce a bias in any study that compares the skills of randomly selected children in OC and TC programs, even when the investigators attempt to control for potentially confounding parameters. The most satisfactory way to address this problem would be to study children assigned to OC or TC programs in a prospective, randomized way. However, this would be very costly, impractical, and many clinicians would contend that such a study would lack clinical equipoise, making it ethically questionable. Thus, like those who have preceded us, we have not attempted to randomize assignment of the children under study to OC or TC mode: we will simply be cautious when interpreting measured differences in speech intelligibility between the two groups. In summary, the goal of this study is to assess the speech intelligibility of children with profound deafness, either congenital or acquired before the age of three, as a function of three independent variables: age, residual hearing, and communication mode.

Table I.

Total number of subjects and total number of data points (in parentheses) in each subject group.

	TC	ORAL	TOTAL
$HA_{90\text{-}100}$	10 (21)	11 (28)	21 (49)
$HA_{100\text{-}110}$	23 (47)	18 (30)	41 (77)
HA_{110+}	41 (64)	23 (29)	64 (93)

Methods

Subjects

We tested 126 profoundly and prelingually deaf children, who were divided into three groups according to their residual hearing, following the classification proposed by Osberger et al. (1993). This classification is based on unaided hearing losses at three frequencies: 500, 1000 and 2000 Hz. $HA_{90\text{-}100}$ is the group of hearing aid users with most residual hearing among the profoundly deaf: they have unaided losses of 90 to 100 dB HL (inclusive). The $HA_{100\text{-}110}$ group includes children with average unaided losses greater than 100 dB HL but less than (or equal to) 110 dB HL. The HA_{110+} group had the least residual hearing, with losses greater than 110 dB HL. All these children were potential candidates for cochlear implantation, and they were tested as part of a longitudinal study of cochlear implant users. Subjects were classified as users of Oral Communication (i.e., their therapy and formal education takes place without the use of signs) or Total Communication, which is the simultaneous use of signs and oral speech. All subjects were in educational programs that emphasized the development of oral skills. Table I shows some characteristics of each group of subjects. Subjects were tested between 1 and 5 times. When a subject was tested more than once, the testing sessions were at least six months apart.

Procedures

Each subject produced 10 sentences. Children under 6 were administered one list from the BIT (Beginner's Intelligibility Test), which uses objects and pictures to convey the target sentence, and an imitative response was elicited after an examiner's spoken model (Osberger, Robbins, Todd, & Riley, 1994). Older children (>6 y.o.) who could read were given the Monsen Sentences Test (Monsen, 1983). Sentences were recorded on cassette tape and digitized. The sentences were played back, in random order, to panels of three listeners with no experience listening to deaf speech The listening sessions were conducted in a double-blind fashion; that is, neither the listeners nor the experimenter knew the communication mode or the amount of residual hearing of the subjects being tested. Listeners heard more than one set of sentences, but these sentences were never from the same list or produced by the same talker. Following Monsen's procedure, each sentence was presented twice and no contextual information was provided (i.e., listeners had no information about the topics related to each sentence). The listeners transcribed what they heard, and their responses were tabulated to determine the percentage of transcribed keywords that were the intended target. Scores for the three judges were then averaged.

Data analysis

The goal of this study was to assess the effect of residual hearing and communication mode on the speech intelligibility of children who have profound, prelingual deafness and who do not use cochlear implants. The effect of residual hearing was analyzed separately for the OC and the TC group, and the effect of communication mode was analyzed separately for each one of the three groups sorted according to residual hearing: $HA_{90\text{-}100}$, $HA_{100\text{-}110}$, and HA_{110+}. To assess the effect of communication mode on speech intelligibility, multivariate nonlinear regressions were done for each one of the three HA groups using the formula:

$$\text{INTELLIGIBILITY} = a_0 + a_1 \times \text{AGE} + a_2 \times \text{MODE} + a_3 \times \text{AGE} \times \text{MODE} \quad [1]$$

MODE refers to communication mode (coded as a discrete variable, with 0 indicating the use of Oral Communication and 1 indicating the use of Total Communication); AGE is the chronological age of each subject when his/her speech was recorded, and INTELLIGIBILITY is the mean intelligibility measured from each recording as explained above. The best-fit regression line for users of Oral Communication (i.e., MODE=0) as a function of age was

$$\text{INTELLIGIBILITY} = a_0 + a_1 \times \text{AGE} \quad [2]$$

and the best-fit regression line for users of Total Communication (i.e., MODE=1) was

$$\text{INTELLIGIBILITY} = (a_0 + a_2) + (a_1 + a_3) \times \text{AGE} \quad [3]$$

When neither a_2 nor a_3 were statistically significant for a particular HA group, this indicated that the regression lines for users of OC and TC were not significantly different. In other words, two separate regression lines (one for OC users and another one for TC users) did not fit the data any better than a single regression line When a_2, a_3, or both were statistically significant, the best-fit lines for the OC and TC groups were examined to see which group had an advantage in speech intelligibility.

The analysis of the effect of residual hearing was conducted in a similar fashion, and it was done separately for the OC and TC groups. Multivariate nonlinear regressions were done using this formula:

$$\text{INTELLIGIBILITY} = (a \times \text{AGE} + b) + (c \times \text{AGE} \times \text{PTA1}) + (d \times \text{AGE} \times \text{PTA 2}) + (e \times \text{PTA1}) + (f \times \text{PTA2}) \quad [4]$$

PTA1 and PTA2 are discrete variables that code the subject's PTA. For group $HA_{90\text{-}100}$, both PTA1 and PTA2 are 1; for group $HA_{100\text{-}110}$ PTA1 is 0 and PTA2 is 1; and for group HA_{110+} both variables are 0. The meaning of the variables AGE and INTELLIGIBILITY has already been explained. The best-fit regression lines for members of each residual hearing group were:

$$\text{INTELLIGIBILITY} = a \times \text{AGE} + b \text{ (for the } HA_{90\text{-}100} \text{ group)}, \quad [5]$$

$$\text{INTELLIGIBILITY} = (a + d) \times \text{AGE} + (b + f) \text{ (for the } HA_{100\text{-}110} \text{ group), and} \quad [6]$$

$$\text{INTELLIGIBILITY} = (a + c + d) \times \text{AGE} + (b + e + f) \text{ (for the } HA_{110+} \text{ group).} \quad [7]$$

The significance of the different regression parameters was examined to determine whether the data were better fit with separate regression lines for each subgroup than with a single line. All the regression analyses were conducted twice: once with all the data points, and another time using only the first data point for each subject (i.e., using a strictly cross-sectional data set). Finally, all the data points were plotted in a single graph for visual inspection of overall trends.

Results

Figure 1 shows speech intelligibility scores as a function of age at testing for children with the least amount of residual hearing: those in the HA_{110+} group. Intelligibility scores were quite low regardless of age. The slopes of the regression lines were very shallow, indicating only small differences as a function of age: 1.2 %/year for TC users and 1.5 %/year for OC users. However, there were significant correlations between speech intelligibility and age, $r = +.64$ for the OC users and $r = +.63$ for TC users ($p < .001$ in both cases). The a_2 and a_3 parameters in the multiple regression were not significant ($p = .77$ and .41, respectively), indicating that the intelligibility data for children in the HA_{110+} group was not fit any better by two regression lines (one for OC users, one for TC users) than by a single line.

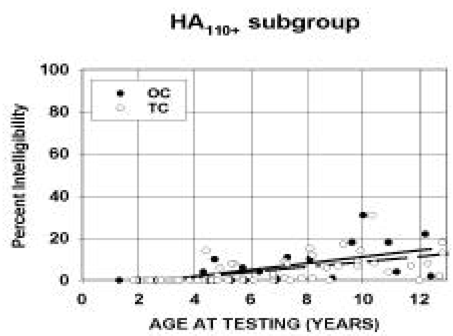

Figure 1. Percent intelligibility as a function of age at testing, for HA_{110+} subjects (PTA>110 dB HL). The black circles represent users of Oral Communication and the black line is the corresponding regression line. Total Communication users are represented by white circles and the corresponding regression line is dashed.

Results were more positive in the $HA_{90\text{-}100}$ group, which included the children with most residual hearing (always within the profoundly hearing impaired range) As Figure 2 shows, there is a very clear trend towards higher intelligibility scores at older ages, both for OC and TC users. The slopes of the regression lines are quite steep, 10.4%/year for the TC group and 7.6%/year for the OC group. The difference between these two slopes was not statistically significant (p=.11 for the a_3 parameter), but the difference between the intercepts of the regression lines for OC and TC users in this $HA_{90\text{-}100}$ group was significant ($p < .001$ for the a_2 parameter). Correlations between speech intelligibility and age were significant ($r = +.81$ for the OC group and $r = +.88$ for the TC group, $p < .001$ in both cases). To the extent that the regression lines are representative of the data, they point to an advantage for the OC group, particularly at the younger ages. All children older than 9.5 years in this sample achieved intelligibility scores higher than 64%. This is a level that would probably allow most of them to have reasonably fluent face-to-face conversations, when the listener can see their face as they talk. In summary, there is a stark contrast between children in the $HA_{90\text{-}100}$ group, who have a relatively good prognosis for their ability to learn how to speak intelligibly (provided that they receive appropriate training) and the children in the HA_{110+} group, whose intelligibility levels are uniformly low.

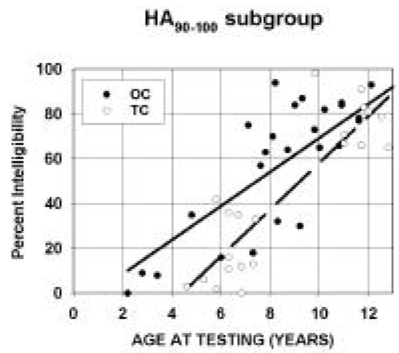

Figure 2. Percent intelligibility as a function of age at testing, for HA_{90-100} subjects (90PTA100 dB HL). The black circles represent users of Oral Communication and the black line is the corresponding regression line. Total Communication users are represented by white circles and the corresponding regression line is dashed.

Results for the group with intermediate amounts of residual hearing, $HA_{100-110}$, are shown in Figure 3. In this group there was a dramatic difference between OC and TC users: the slopes of the regression lines were 7.8%/year for the OC group and only 0.9%/year for the TC group. The coefficients that were significant in the regression were a_2 (p = .002) and a_3 (p < .001), indicating that the speech intelligibility of OC users in this subgroup is significantly different from (and substantially superior to) that of TC users. The regression line for OC users in this group was only slightly lower than the regression lines of OC or TC users in the HA_{90-100} group, and the correlation between intelligibility and age at testing was just as high as for those two subgroups (r = +.86, p < .001). In contrast, the same correlation for TC users in this group was very low (r = +.16) and not significant (p = .30). It is quite apparent that, in contrast to the scores from the other two groups, intelligibility scores for children in the HA100-110 group follow a bimodal distribution. Indeed, scores obtained at ages greater than 9 years old (that is, to the right of the vertical dashed line in Fig. 3), are either greater than 35% or lower than 10%.

Reinforcing this result, Figure 4 shows that there is a trend toward a bimodal distribution for the whole sample of children tested in this study, not just for those in the $HA_{100-110}$ group. The figure shows the particular nature of the data set by joining with thick lines all the data points corresponding to a given subject over time. Some subjects, tested only once, are represented by only one symbol. Some of the children show clearly higher intelligibility scores at higher ages, achieving 50% to 100 % intelligibility by age 10. This group of more successful speakers includes all in the HA_{90-100} group; most OC users and a few TC users in the $HA_{100-110}$ group, and no members of the HA_{110+} group. Other children show very poor intelligibility scores regardless of age and in spite of all the speech training that they receive. Their scores rarely exceed 30% and are typically lower than 20%. These children include all members of the HA_{110+} group; a few OC users and most TC users in the $HA_{100-110}$ group, and no members of the HA_{110+} group.

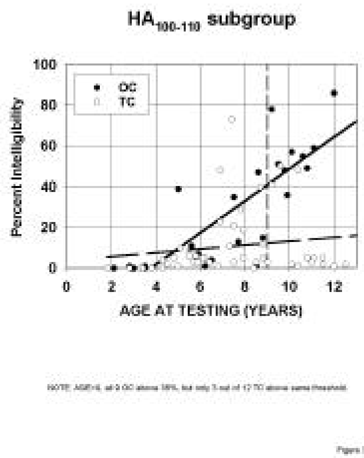

Figure 3. Percent intelligibility as a function of age at testing, for $HA_{100\text{-}110}$ subjects (100 < PTA110 dB HL). The black circles represent users of Oral Communication and the black line is the corresponding regression line. Total Communication users are represented by white circles and the corresponding regression line is dashed. The vertical dashed line at age 9 helps visualize the bimodal distribution observed at older ages in this subject group.

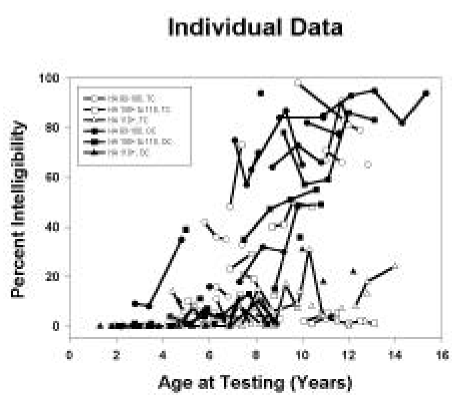

Figure 4. All the individual data points for OC and TC users within each subgroup. Successive data points for the same subject are connected by solid lines.

The careful reader will not be surprised to read that for both the OC and the TC groups, data were significantly better fit by three regression lines (one for each residual hearing subgroup) than by a single line[3]. In both groups, there was a significant trend toward higher intelligibility scores for those subgroups with greater amounts of residual hearing.

Discussion

The data show two strong overall trends, neither of which should be surprising: children with more residual hearing tend to be more intelligible, and some subgroups of OC users tend to be more intelligible than their TC-using counterparts. Although this second conclusion should be tempered due to the considerations discussed in the introduction (namely, that assignment to the OC or TC methods is not a random process), it may still be quite enlightening to ponder the differences between OC and TC users in each one of the subgroups. These differences were quite dramatic in the $HA_{100\text{-}110}$ subgroup, much less so in the $HA_{90\text{-}100}$ subgroup, and nonexistent in the HA_{110+} subgroup. It might be argued that a floor effect decreased the power of the comparison in this last subgroup, but it is still clear that the intelligibility levels of all children in the subgroup, OC and TC users alike, are functionally quite low and may not allow easy communication with naive listeners. Remember that a 20% intelligibility level means that a naive listener would understand one word out of every five uttered by the speaker. Even in a situation with high levels of contextual information, a listener would have a hard time communicating when only one word out of five is identified reliably. The low levels of intelligibility in this subgroup raise the question whether the many hours that these children have spent in speech training will pay off. It is important to remember that this consideration applies only to children with PTAs greater than 110 dB and who have not received cochlear implants, because cochlear implants typically increase the child's auditory skills to a level where he or she might derive much more benefit from oral rehabilitation than the typical child in the HA_{110+} subgroup who uses hearing aids.

The $HA_{90\text{-}100}$ subgroup is, in a way, the mirror image of the HA_{110+} subgroup. Again, the differences between OC and TC users are not overwhelming, but unlike in the HA_{110+} subgroup, this happens because both OC and TC users show much improved speech intelligibility at older ages. It is in the $HA_{100\text{-}110}$ subgroup that the average differences between OC and TC users are truly remarkable. Perhaps there is a particular range of hearing impairment where an intensive concentration on oral rehabilitation results in maximum payoff. The present results suggest that such a range may be in the vicinity of 100 to 110 dB HL, at least from the point of view of development of intelligible speech.

In contrast to the intricate pattern of differences between OC and TC users found in the present study, Geers and Moog (1992) found a clear advantage for OC users over TC users for each subgroup of children, regardless of the amount of residual hearing. However, there are a number of differences between the two studies that make a direct comparison of the results difficult. Firstly, Geers and Moog used the SPINE test instead of the write down procedure employed in this study. One consequence of using SPINE is that the listeners in that study were aware of the child's identity and communication mode (i.e., the listeners were not "blind" to the independent variable of communication mode). Secondly, The Geers and Moog study included speakers who were 16 or 17 years old, much older than the subjects in this study. The age difference might explain some of the differences between the two studies. For example, it's possible that the OC users in the HA_{110+} group in our study might outperform the speech intelligibility levels of their TC counterparts in future years. However, other differences between the two studies are more difficult to explain. In particular, most of the older TC subjects in our $HA_{90\text{-}100}$ subgroup

[3] The coefficients that were statistically significant in the OC regression were: d ($p < .001$), e ($p = .031$) and f ($p = .004$). The coefficient a was marginally nonsignificant at $p = .057$. In the TC regression, the significant coefficients were: a ($p = .010$), c ($p < .001$), and e ($p < .001$).

(those who were 9 to 12 years old) scored higher than Geers and Moog's TC subjects who had the same amount of residual hearing. This happened even though the Geers and Moog subjects were older and the listeners in that study had an easier task than in ours, namely, they only had to select one word out of four possible ones instead of writing down what the speaker said. One possibility is that the TC subjects in both studies were not similar. For example, our TC users may have received more oral training than those in the Geers and Moog study.

Finally, and even though the study of children with cochlear implants is not a focus of the present study, it must be said that children who have received implants before the age of six and who have used state-of-the-art devices and stimulation strategies since initial stimulation seem to perform at least as well as their hearing aid using peers in the $HA_{90\text{-}100}$ group, from the viewpoints of speech perception (Svirsky & Meyer, 1999; Meyer & Svirsky, in press), speech production (Svirsky et al., in press a) and language development (Svirsky, in press; Svirsky et al. in press b). Thus, the prognosis for oral communication for children who use cochlear implant technology may be quite more optimistic than those for any of the subgroups that are analyzed in this study. It is our hope that the present study will serve as a benchmark to refine our comparisons of speech intelligibility by cochlear implant and hearing aid users. Such comparisons, together with parallel studies of speech perception and language development, will continue to strengthen the knowledge base that we draw upon when we need to make clinical decisions about pediatric cochlear implantation.

References

Becker, R., Schildhammer, A., & Ruoß, M. (1994). Sprachliche Leistungen von gehörlosen Kindern und Jugendlichen beim mündlichen und schriftlichen Erzählen einer Bildergeschichte. *Zeitschrift für experimentelle und angewandte Psychologie*, 61, 349-377.

Brannon, J. B. (1964). *Visual feedback of glossal motions and its influence upon the speech of deaf children.* Unpublished doctoral dissertation, Northwestern University, Evanston, IL.

Caccamise, F., Hatfield, N., & Brewer, L. (1978) Manual/simultaneous communication (M/SC) research: results and implications. *American Annals of the Deaf* (Silver Spring, MD). 123(7):803-23.

Cowan, R.S.C., Brown, C.J., Whitford, L.A., et al. (1995). Speech perception in children using the advanced SPEAK speech-processing strategy. *Annals of Otology, Rhinology, & Laryngology, 104*, 318-321.

Fairbanks, G. (1960). *Voice and articulation drillbook*. (Second ed.). New York: Harper & Row.

Geers, A. E., & Moog, J. S. (1992). Speech perception and production skills of students with impaired hearing from oral and total communication education settings. *Journal of Speech and Hearing Research*, 35, 1384-1393.

Gold, T. (1980). Speech production in hearing-impaired children. *Journal of Communication Disorders*, 13, 397-418.

Guilder, R. P., & Hopkins, L. A. (1936). The importance of auditory functions studies in the educational program for the auditorily handicapped child. *The Volta Review*, 38, 69-74, 149-155.

Hippocrates. (1853). *Œuvres complètes d'Hippocrate* (Vol. 8) (E. Littré, Ed. and Trans.). Paris: J. B. Baillière.

Hirsch, I. J., Davis, H., Silverman, S. R., Reynolds, E. G., Eldert, E., & Benson, R. W. (1952). Development of materials for speech audiometry. *Journal of Speech and Hearing Disorders*, 17, 321-337.

Hudgins, C. V., & Numbers, F. C. (1942). An investigation of the intelligibility of the speech of the deaf. *Genetic Psychology Monographs*, 25, 289-392.

John, J. E. J., & Howarth, J. N. (1965). The effect of time distortions on the intelligibility of deaf children's speech. *Language and Speech*, 8, 127-134.

Maasen, B., & Povel, D.-J. (1985). The effect of segmental and suprasegmental corrections on the intelligibility of deaf speech. *Journal of the Acoustical Society of America*, 78, 877-886.

Markides, A. (1970). The speech of deaf and partially-hearing children with special reference to factors affecting intelligibility. *British Journal of Disorders of Communication*, 5, 126-140.

McGarr, N. S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech and Hearing Research*, 26, 451-458.

Metz, D. E., Schiavetti, N., & Sitler, R. W. (1980). Toward an objective description of the dependent and independent variables associated with intelligibility assessments of hearing-impaired adults. In J. D. Subtelny (Ed.), *Speech assessment and speech improvement for the hearing impaired* (pp. 72-81). Washington, DC: Alexander Graham Bell Association for the Deaf.

Meyer, T.A. & Svirsky, M.A. (In press). Speech perception by children with the Clarion (CIS) or Nucleus (SPEAK) cochlear implant or hearing aids. *Annals of Otology, Rhinology, and Laryngology*.

Miyamoto, R.T., Kirk, K.I., Robbins, A.M., Todd, S.L. & Riley, A.I. (1996). Speech perception and speech production skills of children with multichannel cochlear implants. *Acta Oto-Laryngologica, 116,* 240-243.

Monsen, R. B. (1978). Toward measuring how well hearing-impaired children can speak. *Journal of Speech and Hearing Research*, 21, 197-219.

Monsen, R. B. (1981). A usable test for the speech intelligibility of deaf talkers. *American Annals of the Deaf*, 126, 845-852.

Monsen, R. B. (1983). The oral speech intelligibility of hearing-impaired talkers. *Journal of Speech and Hearing Disorders*, 48, 286-296.

Osberger, M. J., & Levitt, H. (1979). The effect of timing errors on the intelligibility of deaf children's speech. *Journal of the Acoustical Society of America*, 66, 1316-1324.

Osberger, M. J., Maso, M., & Sam, L. K. (1993). Speech intelligibility of children with cochlear implants, tactile aids, or hearing aids. *Journal of Speech and Hearing Research*, 36, 186-203.

Osberger, M. J., Robbins, A. M., Todd, S. L,, & Riley, A. I. (1994). Speech intelligibility of children with cochlear implants. *The Volta Review*, 96, 169-180.

Osberger, M.J., Robbins, A.M. Todd, S.L., Riley, A.I., & Miyamoto, R.T. (1994). Speech production skills of children with multichannel cochlear implants. In I.J. Hockmair-Desoyer & E.S. Hochmair (Eds.), *Advances in cochlear implants* (pp. 503-508). Vienna: Manz.

Robbins, A.M., Kirk, K.I., Osberger, M.J. & Ertmer, D.J. (1995). Speech intelligibility of implanted children. *Annals of Otology, Rhinology, and Laryngology, 104,* 399-401.

Samar, V. J., & Metz, D. E. (1988). Criterion validity of speech intelligibility rating-scale procedures for the hearing-impaired population. *Journal of Speech and Hearing Research*, 31, 307-316.

Skinner, M.W., Clark, G.M., Whitford, L.A. et al., (1994). Evaluation of a new spectral peak (SPEAK) strategy for the Nucleus 22 channel cochlear implant system. *American Journal of Otology, 15,* 15-27.

Smith, C. R. (1972). *Residual hearing and speech production in deaf children*. Unpublished doctoral dissertation, City University of New York, New York, NY.

Smith, C.R. (1973). Residual hearing and speech production in deaf children. *CUNY Communications Science Laboratory Research Report, Volume 4*.

Smith, C. R. (1975). Residual hearing and speech production in deaf children. *Journal of Speech and Hearing Research*, 18, 795-811.

Smith, C. R. (1982). Differences between read and spontaneous speech of deaf children. *Journal of the Acoustical Society of America*, 72, 1304-1306.

Subtelny, J. D. (1977). Assessment of speech with implications for training. In F. H. Bess (Ed.), *Childhood deafness: Causation, assessment, and management* (pp. 183-194). New York: Grune & Stratton.

Svirsky, M.A., & Meyer, T.A.. (1999). A comparison of speech perception for pediatric Clarion and hearing aid users. *Annals of Otology, Rhinology, and Laryngology*, 108 (Suppl 177), 104-110.

Svirsky, M.A., Sloan, R.B., Caldwell, M., & Miyamoto, R.T. (In press a). Speech intelligibility of prelingually deaf children with multichannel cochlear implants. *Annals of Otology, Rhinology, and Laryngology*.

Svirsky, M.A., Robbins, A.M., Kirk, K.I., Pisoni, D.B., & Miyamoto, R.T. (In press b). Language development in profoundly deaf children with cochlear implants. *Psychological Science*.

Svirsky, M.A. (In press). Language development in children with profound and prelingual hearing loss, without cochlear implants. *Annals of Otology, Rhinology, and Laryngology*.

Wilbur, R.B. (1979). *American Sign Language and sign systems*. Baltimore, MD: University Park Press.