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

Audio-Visual Integrative Abilities of Prelingually Deafened Children with Cochlear Implants: A First Report[1]

Lorin Lachs, Karen I. Kirk,[2] and David B. Pisoni[2]

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was funded, in part, by NIH-NIDCD Grants R01 DC00111, T32 DC00012, K08 DC00126, and K08 DC00083.
[2] DeVault Otologic Research Laboratory, Department of Otolaryngology-Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, IN.

Audiovisual Integrative Abilities of Prelingually Deafened Children with Cochlear Implants: A First Report

Abstract. Although there has been a great deal of recent empirical work and new theoretical interest in audio-visual perception of speech in both normal-hearing and hearing-impaired adults, relatively little is known about the development of these abilities and skills in deaf children with cochlear implants (CIs). This study examined the way in which children integrate visual speech cues available in the talker's face with auditory speech cues provided by their CIs to enhance spoken language comprehension. A measure of audiovisual integration ability, R, developed by Sumby and Pollack (1954), was computed using sentence comprehension scores obtained from the Common Phrases Test that was administered using auditory and audiovisual conditions. R represents the amount of gain provided in the AV condition relative to the possible gain in performance in the A condition. Because R is normalized relative to absolute performance in the A condition, R measures the ability of a perceiver to utilize and integrate combined auditory and visual information. The results of our analyses indicated that children who were better at recognizing isolated words through listening alone were also better at combining and integrating the complementary information about articulation that is available under multimodal presentation. In addition, we found that children who were better integrators also displayed higher speech intelligibility scores. Treatment programs that aim to increase receptive or productive ability in children who have CIs, therefore, may wish to focus and emphasize the inherent cross-correlations between auditory and visual information in speech.

Introduction

With continued broadening of candidacy criteria and technological advances in cochlear implant (CI) signal processing, more children than ever before have the potential to develop spoken word recognition and language processing skills through a CI. However, there are enormous individual differences in pediatric CI outcomes (Fryauf-Bertschy, Tyler, Kelsay, & Gantz, 1992; Fryauf-Bertschy, Tyler, Kelsay, Gantz, & Woodworth, 1997; Miyamoto et al., 1989; Osberger et al., 1991; Staller, Beiter, Brimacombe, Mecklenberg, & Arndt, 1991; Tyler et al., 1997a; Tyler, Fryauf-Bertschy, Gantz, Kelsay, & Woodworth, 1997b; Tyler et al., 1997c; Zimmerman-Phillips, Osberger, & Robbins, 1997). Some children can communicate extremely well using the auditory/oral modality and acquire age-appropriate language skills, whereas other children may display only minimal spoken word recognition skills and/or display very delayed language abilities (Bollard, Chute, Popp, & Parisier, 1999; Pisoni, Svirsky, Kirk, & Miyamoto, submitted; Robbins, Bollard, & Green, 1999; Robbins, Svirsky, & Kirk, 1997; Svirsky, Sloan, Caldwell, & Miyamoto, 1998; Tyler, Tomblin, Spencer, Kelsay, & Fryauf-Bertschy, in press). These differences currently cannot be predicted prior to implantation but emerge only after electrical stimulation has commenced. Although demographic and audiologic factors play an important role in pediatric CI outcomes, these variables alone are not sufficient to account for the large individual differences (Fryauf-Bertschy et al., 1992; Fryauf-Bertschy et al., 1997). Rather, variability in performance may be related to perceptual, cognitive and linguistic processes involved in the acquisition and processing of spoken language by the central nervous system (Pisoni & Geers, 1998).

One important perceptual process that may provide some insight into the basis of individual differences in multimodal sensory integration. It is well known that normal-hearing adults can make use of both auditory and visual information to enhance speech intelligibility by up to +15 dB (Massaro & Cohen, 1995; Rosenblum, Johnson, & Saldaña, 1996; Sumby & Pollack, 1954; Summerfield, 1987). Relatively little, however, is known about the integrative abilities of prelingually deaf children with CIs.

Cross-modal integration is thought to reflect the ability to use different sources of information in order to perceive the underlying articulations of the talker's vocal tract (Fowler, 1989; Fowler & Dekle, 1991; Massaro & Cohen, 1995; Remez, Fellowes, Pisoni, Goh, & Rubin, 1999; Rosenblum, 1994; Rosenblum & Saldaña, 1996; Vatikiotis-Bateson, Munhall, Kasahara, Garcia, & Yehia, 1996). The seminal work in this area was conducted by Sumby and Pollack (1954) who showed that the intelligibility of spoken words in noise could be increased by as much as +15 dB when normal-hearing participants were able to both hear and see the speaker uttering test items. Because this gain is so large, perceptual integration has the potential to be extremely useful for the study and treatment of speech perception in individuals with hearing impairments. The findings on cross-modal integration may also provide new information about the basis of individual differences among children with cochlear implants.

Some analysis of the integrative abilities of individuals with hearing impairments has been carried out (see Massaro & Cohen, 1999 for a meta-analysis and modeling of some of these studies). For example, Erber (1972) tested the speech perception abilities of children with normal hearing, severely impaired hearing, and profoundly impaired hearing under auditory-alone, visual-alone and auditory-visual conditions. The stimuli used in this experiment were consonants embedded in an /aCa/ environment. Like the participants in Sumby and Pollack's (1954) study, Erber found that the hearing impaired children were able to capitalize on the complementary information provided by the visual modality in the auditory-visual condition, increasing their speech perception performance relative to their performance in the auditory-alone condition. In another study, consonant confusions made by two multiple-channel cochlear implant users were explored when stimuli were presented in auditory-alone, visual-alone, and auditory-visual conditions (Dowell et al., 1982). They found that speech intelligibility was enhanced when patients were able to combine auditory information in the form of electrical stimulation through the CI and visual information. Taken together, these studies demonstrate the nearly universal finding that audiovisual information facilitates speech perception performance relative to performance with auditory-only or visual-only information.

Substantial variation, however, exists in the extent to which an individual's scores are facilitated. Grant, Walden and Seitz (1998) compared observed audiovisual facilitation to that predicted by several models of integration. These models predicted performance based on the unimodal recognition scores for segmental identification. From these unimodal scores, measures of optimal audiovisual integration can be calculated. Grant et al. found that not all individuals integrate auditory and visual information optimally, especially when speech perception accuracy was measured using higher-order units of speech like words and sentences. While the predictions based on segmental accuracy could account for a large amount of the variance observed (e.g., 50% of sentence-level integration variability), the authors conclude that much more work must be carried out before an adequate model of word- and sentence-level integration can be formed. They speculate that such a model will incorporate both lexical factors and semantic contextual information.

Lexical factors have been shown to play a role in the variability observed in the speech perception of cochlear implant users. The Neighborhood Activation Model of spoken word recognition (Luce, 1986; Luce & Pisoni, 1998) proposes that all spoken words are perceived and recognized within the context of similar words contained in the mental lexicon, the mental storehouse of words known by a listener. In this model, three factors are important for the recognition of words. First, the frequency of occurrence of a word in the language acts as a bias for word recognition, such that more frequent words will be recognized more easily than less frequent words. Second, the model assumes that all words with similar acoustic/phonetic form (called the “neighbors”) compete with each other and the target word for activation during the word recognition process. Accordingly, words from dense neighborhoods are less easily recognized than words from sparse neighborhoods. Finally, because the frequency bias acts on the neighbors as well as the target word, the average frequency of occurrence in the neighborhood plays a role in accuracy of recognition, as well. If a target word is of low frequency relative to the frequency in the neighborhood, then it will be harder to recognize than if it were of high frequency relative to the rest of the neighborhood.

These factors have been incorporated into two tests of open set spoken language processing ability. The Lexical Neighborhood Test (LNT) and multisyllabic Lexical Neighborhood Test (MLNT, Kirk, 1999; Kirk, Eisenberger, Martinez, & Hay-McCutcheon, 1999; Kirk, Pisoni, & Osberger, 1995) incorporate words from two extreme combinations of the lexical factors mentioned above. Lexically “easy” words are words from sparse similarity neighborhoods with relatively high frequencies of occurrence. Lexically “hard” words are words from dense similarity neighborhoods with relatively low frequencies. These lexical factors were calculated relative to children's productive vocabulary (Kirk, 1999; Logan, 1992). Initial analysis of performance on these two tests by children with cochlear implants showed that word recognition accuracy on lists of easy words was better than accuracy on lists of hard words, indicating that children with cochlear implants recognize words in the context of the other words contained in their lexicons (Kirk et al., 1995).

Given the fact that both the variation in audiovisual integrative ability and the variation in cochlear implant users' auditory-alone speech perception ability seem to relate to an underlying level of lexical representation, the present study was carried out to investigate the relationship between auditory-alone measures of spoken language comprehension and audiovisual integrative abilities in hearing-impaired children who use cochlear implants.

Method

Participants

Participants were 27 children with prelingual deafness who had used a multichannel cochlear implant for two years. The average age of onset of deafness was 0.51 years. Their average age at implantation was 4.52 years. The mean unaided auditory threshold (as measured with pure tones) was 112.20 dB HL. Table 1 shows the specific demographic data for each of the participants involved.

Procedures and Measures

Three tests designed to measure spoken word recognition performance in audio-alone conditions were administered live-voice to participants by an audiologist or speech-language pathologist. The Lexical Neighborhood Test (LNT) and Multisyllabic Lexical Neighborhood Test (MLNT, Kirk, 1999; Kirk et al., 1999; Kirk et al., 1995) are new open-set tests of word recognition for children that assess the effects of word frequency and lexical similarity on spoken word recognition in children. Lexical similarity is measured by calculating the number of words that differ from the target word by only one phoneme. Words from “dense” neighborhoods (i.e., words with many neighbors) tend to be more difficult to identify than words that come from “sparse” neighborhoods. Additionally, high frequency words tend to be recognized more easily than low frequency words. Using these lexical factors, it is possible to classify the items on the LNT and MLNT tests into “Easy” and “Hard”. Easy words are high frequency words that reside in relatively low frequency, sparse neighborhoods. Hard words are low frequency words that reside in relatively high frequency, dense neighborhoods. Differences between easy and hard words will be maintained throughout the rest of this report because they provide important information about how lexical competition operates during word recognition.

Participants	Etiology	Age at profound loss (yr.)	Unaided PTA (dB HL)	Age CI fit (yr.)	Processor	Strategy	# Active Electrodes	Age at testing (yr.)
1	unknown	Congenital	116.73	3.50	MSP	MPEAK	18.00	5.50
2	unknown	Congenital	111.67	5.50	MSP	MPEAK	22.00	7.70
3	unknown	1.00	120.07	4.90	MSP	MPEAK	22.00	6.80
4	meningitis	.40	111.67	3.80	SPECTRA	SPEAK	18.00	5.90
5	unknown	Congenital	101.67	5.80	MSP	MPEAK	13.00	7.80
6	unknown	Congenital	116.73	4.10	MSP	MPEAK	19.00	6.10
7	genetic	Congenital	103.33	5.20	MSP	MPEAK	22.00	6.90
8	unknown	Congenital	118.43	4.90	WSP	F...	13.00	6.90
9	genetic	3.00	113.37	5.20	MSP	MPEAK	19.00	7.30
10	unknown	Congenital	110.00	3.70	SPECTRA	SPEAK	18.00	5.90
11	meningitis	1.90	118.40	5.40	MSP	F...	8.00	8.00
12	unknown	Congenital	108.37	4.40	SPECTRA	SPEAK	18.00	6.30
13	unknown	Congenital	108.37	4.60	MSP	MPEAK	18.00	6.60
14	genetic	Congenital	111.70	5.30	SPECTRA	SPEAK	19.00	7.20
15	unknown	Congenital	105.00	5.00	MSP	MPEAK	19.00	7.00
16	unknown	Congenital	118.40	5.30	SPECTRA	SPEAK	20.00	7.30
17	unknown	Congenital	100.00	5.20	SPECTRA	SPEAK	22.00	7.40
18	unknown	Congenital	111.67	4.30	MSP	MPEAK	22.00	6.30
19	meningitis	1.80	116.73	4.30	WSP	F...	15.00	6.30
20	meningitis	1.80	116.73	5.30	MSP	MPEAK	16.00	7.10
21	unknown	Congenital	98.33	4.30	SPECTRA	SPEAK	20.00	6.30
22	meningitis	1.40	113.40	2.20	SPECTRA	SPEAK	18.00	4.20
23	unknown	Congenital	113.37	2.90	MSP	F...	21.00	4.90
24	meningitis	1.20	115.07	4.60	MSP	MPEAK	19.00	6.60
25	genetic	Congenital	118.40	4.20	SPECTRA	SPEAK	20.00	6.30
26	meningitis	1.30	113.40	3.10	MSP	MPEAK	19.00	5.30
27	unknown	Congenital	118.43	5.00	MSP	MPEAK	19.00	7.10
Means		0.51	112.20	4.52			18.41	6.56

Table 1. The demographic characteristics of the participants in the present study. For all participants, the length of device use was 2 years.

In addition to the LNT and MLNT tests, the Phonetically Balanced Kindergarten word lists (PB-K, Haskins, 1949) were administered via live voice to test speech perception in an audio-only setting. This test is an open-set measure whose items are balanced for phonetic content. The PB-K is a widely used measure of speech perception ability in children who use cochlear implants (Kirk, Diefendorf, Pisoni, & Robbins, 1997; Meyer & Pisoni, 1999). The three measures referred to above (LNT, MLNT and PB-K) will herein be referred to collectively as the “auditory measures.”

A fourth test, the Common Phrases (CP) Test, was administered under three conditions, auditory-alone (A), visual-alone (V) and audiovisual (AV). In this report, we only focus on data for the A and AV conditions because we were interested in the possible gain in performance from visual integration. This test measures the ability to understand phrases used in everyday situations, like “It is cold outside”. Performance in each condition is scored by the percentage of phrases correctly repeated by the child. The scores in each of these conditions can be combined to measure R, the relative gain in auditory intelligibility due to the availability of visual information about articulation (Sumby & Pollack, 1954).

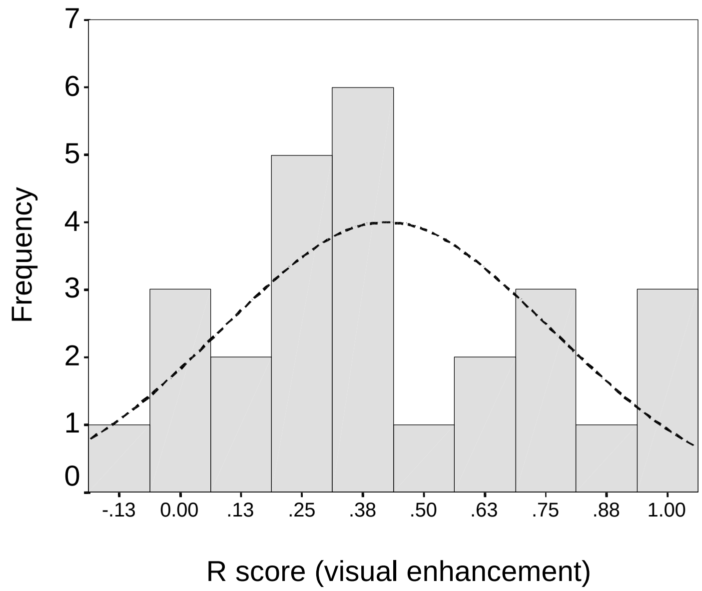

Figure 1. Frequency distribution of R (visual enhancement) scores for the sample. Higher R scores denote higher gains in accuracy in the audiovisual condition relative to accuracy in the audio-alone condition.

R is computed using the following formula:

$$R = {AV - A}\big/{100 - A}$$

where AV and A represent the accuracy scores obtained in the audiovisual and auditory-alone conditions. From the formula, one can see that R describes the gain in accuracy in the AV condition relative to the accuracy in the A condition, normalized relative to the amount by which intelligibility could have possibly improved. We take R to operationally define the ability to integrate information from disparate sensory modalities.

Finally, a test of spoken language production was administered for each child (Miyamoto et al., 1997). In this test, ten sentences produced by the child are recorded and played back to 3 listeners for transcription. Intelligibility is measured by calculating the average number of words correct for the three talkers.

Results

Audiovisual Integration Scores

Figure 1 shows the frequency distribution of R scores for children in the sample under study in the present experiment. The normal curve superimposed over the data is based on the mean and standard deviation (*M* = 0.42, *SD* = 0.34) of the sample. As shown in the figure, our sample was relatively normally distributed with respect to audiovisual integrative ability, with a slight positive skew. It is also interesting to note that these scores vary over a very large range. Some children were able to maximally

Participant	Modality of Common Phrases Auditory-alone	Visual-alone	Audiovisual	R score
1	.00	20.00	30.00	.30
2	20.00	50.00	50.00	.38
3	.00	.00	.00	.00
4	80.00	50.00	100.00	1.00
5	80.00	80.00	100.00	1.00
6	.00	30.00	40.00	.40
7	.00	40.00	20.00	.20
8	10.00	.	80.00	.78
9	60.00	.	70.00	.25
10	80.00	20.00	90.00	.50
11	.00	80.00	90.00	.90
12	70.00	50.00	90.00	.67
13	.00	20.00	10.00	.10
14	90.00	50.00	100.00	1.00
15	10.00	.00	30.00	.22
16	20.00	10.00	20.00	.00
17	30.00	40.00	60.00	.43
18	20.00	.00	20.00	.00
19	40.00	.	60.00	.33
20	.00	50.00	60.00	.60
21	70.00	60.00	80.00	.33
22	40.00	.00	30.00	-.17
23	.00	.	40.00	.40
24	10.00	.	80.00	.78
25	20.00	.	30.00	.13
26	.00	.00	20.00	.20
27	.00	.	70.00	.70
Mean	27.7778	32.5000	54.4444	.4231

Table 2. Performance for each participant on each of the subtests of the Common Phrases test, along with the R score calculated from those measures. A "." indicates that the participant was not tested under that condition.

capitalize on the additional visual information in the AV condition, indicated by R scores of 1.0, while others received little if any benefit, indicated by the zero and negative values. Table 2 shows the specific scores obtained by each child in the auditory-alone, visual-alone, and audiovisual conditions of the Common Phrases test, along with the corresponding R scores. Clearly, children with cochlear implants exhibit a wide range of multimodal integrative abilities in this task.

Correlations with Word Recognition Scores

We were also interested in whether measures of spoken word recognition were related to cross-modal integration. As a first pass at answering this question, the median score for each auditory measure was calculated, and participants were categorized as being in either the low or high group of the split. Using this method, it was entirely possible for a particular patient to be classified in the low group for the split based on one measure and be classified in the high group for the split based on another measure. However, when this happened for a particular child, it only affected classification based on one test. Because not all of the children participated in each of the auditory-alone measures, the total Ns for the various splits differed. The median score, along with the N, for each auditory measure is displayed in Table 3. The maximum score obtainable on each of these measures was 100.

	Auditory Speech Perception Test				
	LNT Easy	**LNT Hard**	**MLNT Easy**	**MLNT Hard**	**PB-K**
Median	28	20	47	53	8
N	17	13	15	11	24

Table 3. Median values and Ns for the five measures of auditory perception.

Figure 2 shows the R scores for the low and high median split groups. Each panel shows the results based on a different median split. The shaded bar in each panel shows the average R score for children in the low group of the median split, and the white bar shows the average R score for children in the high group of the median split. Overall, there is a consistent numerical trend for children in the high median split group to also have higher R scores. However, for some median splits, this trend is more evident than it is for others.

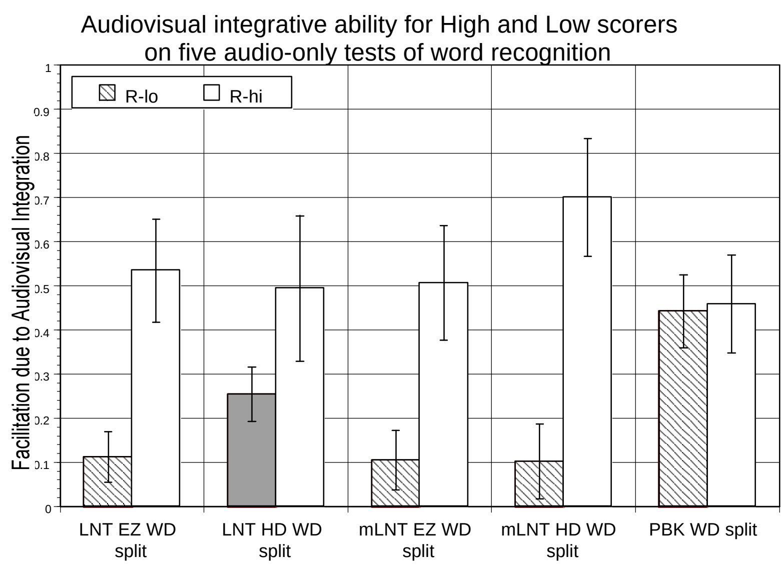

Figure 2. Audiovisual integrative ability (R) for the High and Low groups of the median splits for five measures of auditory-alone word recognition. Error bars are standard errors.

Two-tailed, independent sample t-tests were conducted (using an α level of 0.05) to determine if there were differences in the R scores for children in the high scoring vs. low scoring groups for each auditory measure. Significant differences were found when the children were split by median scores on the LNT Easy test, $t(15) = 3.36$, $p = 0.004$, the MLNT Easy test, $t(13) = 2.62$, $p = 0.021$, and the MLNT Hard test, $t(9) = 3.927$, $p = 0.003$. However, the statistics failed to show a significant difference for splits based on the LNT Hard test, $t(11) = 1.275$, *ns*, and the PB-K test, $t(22) = 0.116$, *ns*. As shown below, these were the two most difficult measures. The data from these t-tests indicates that children who were good performers on auditory-alone measures of spoken word recognition were also good integrators.

In order to more fully characterize the relationship between audiovisual integrative ability and our measures of speech perception, bivariate correlations were calculated between R scores and each of the three auditory measures. Table 4 shows the correlations between each of the measures and R. The correlations show the same pattern of results as the t-tests using median splits. There was a significant relationship between the R score and performance on the LNT Easy, MLNT Easy, and MLNT Hard tests. However, there was no significant relationship between R scores and performance on the LNT Hard and PB-K tests.

	Auditory Speech Perception Test				
	LNT Easy	**LNT Hard**	**MLNT Easy**	**MLNT Hard**	**PB-K**
R score	0.78**	0.28	0.57*	0.68*	0.28

Table 4. The correlation (r) between R scores and the various measures of speech perception. ** indicates significance using an α-level of 0.01; * indicates significance using an α-level of 0.05.

Figure 3 shows the scatterplots for each of these conditions. In each figure, the Y-axis represents the R score and the X-axis represents the score on one of the spoken word recognition tests. Each triangle represents one participant. Shaded triangles are for participants who scored below the median for the test represented on the X-axis, open triangles are for participants who scored above the median. These scatterplots also show the regression line for each relationship. Not surprisingly, the graphs show a strong relationship between R score and the LNT Easy test and both subsections of the MLNT test.

An examination of the scatterplots shows that the lack of a significant relationship between R and the LNT Hard score may be due to an unusually low outlier (the shaded triangle with a negative R score and an LNT Hard score of around 33). This participant is below the median score for all the other tests, except the PB-K, which also failed to show a significant relationship with R. In both the LNT Hard test and the PB-K test, the range of scores is reduced relative to the other tests. Consequently, the median score is relatively low. Table 5 shows the skew values associated with the distribution of scores for each of the auditory measures. Examination of the table shows that the distribution of scores for the LNT Hard and PB-K tests are more positively skewed than the distributions for the other measures used in the present study. These distributional measures confirm that the LNT Hard and PB-K tests were extremely difficult for most of the children involved in our study. We suspect that the lack of significant correlations between these tests and R are due to a floor effect.

	Auditory Speech Perception Test				
	LNT Easy	**LNT Hard**	**MLNT Easy**	**MLNT Hard**	**PB-K**
Skewness	.377	.890	0.322	-0.655	1.167

Table 5. Measures of skewness for the five auditory measures of word recognition. Positive values denote rightward skews and negative values denote leftward skews.

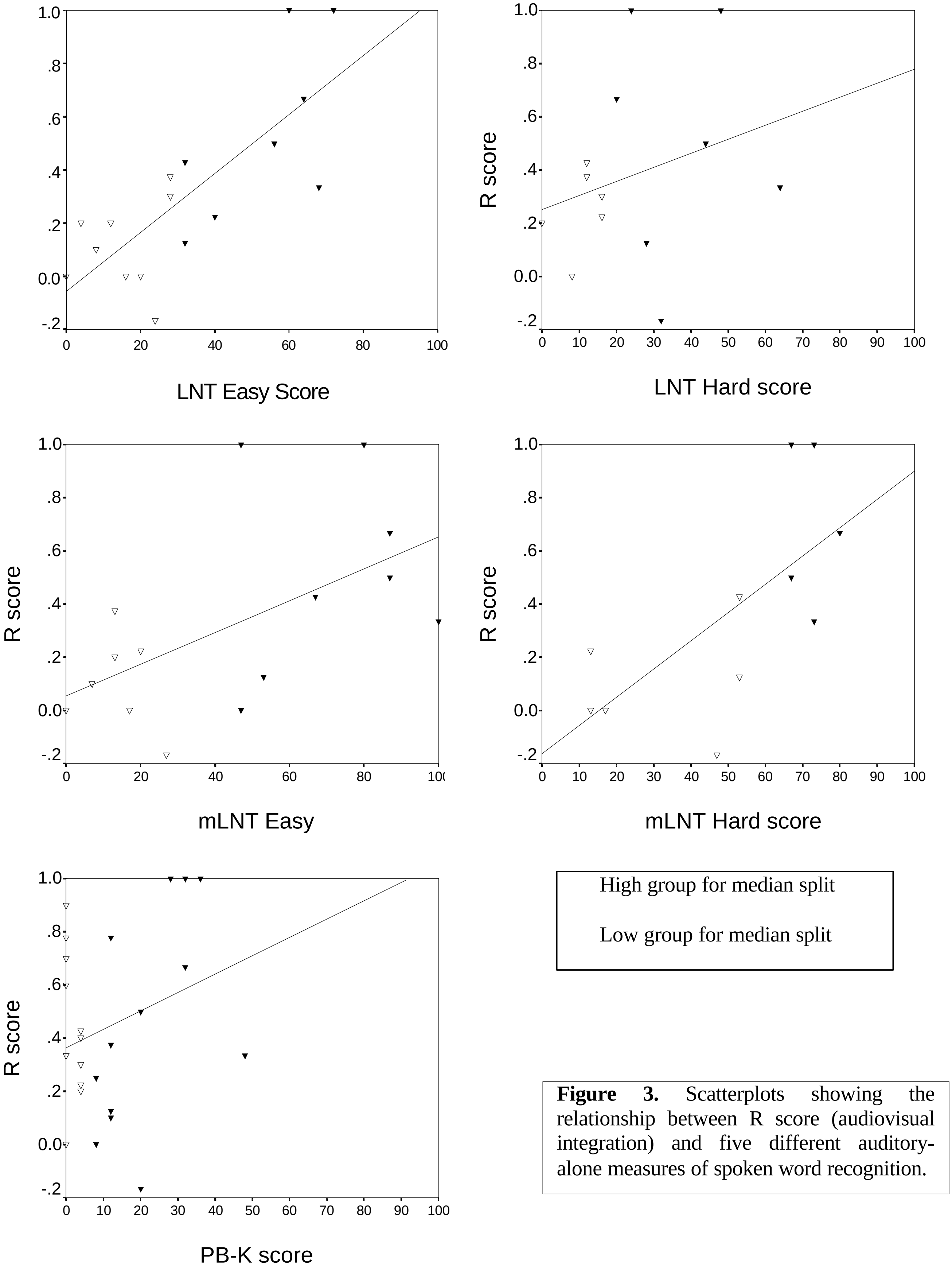

Figure 3. Scatterplots showing the relationship between R score (audiovisual integration) and five different auditory-alone measures of spoken word recognition.

Speech Intelligibility

In addition to the word recognition scores, 23 of the 27 participants also provided scores on a test of speech intelligibility. The mean intelligibility score was 17.13 with a standard deviation of 12.47. The range of intelligibility scores spanned from 2% to 45%. To assess the relations between R and speech intelligibility, we carried out a correlation. The results showed that there was also a significant correlation between intelligibility score and R score, $r(23) = +0.421$, $p = 0.046$. This relationship denotes that children with more intelligible speech are also better audiovisual integrators.

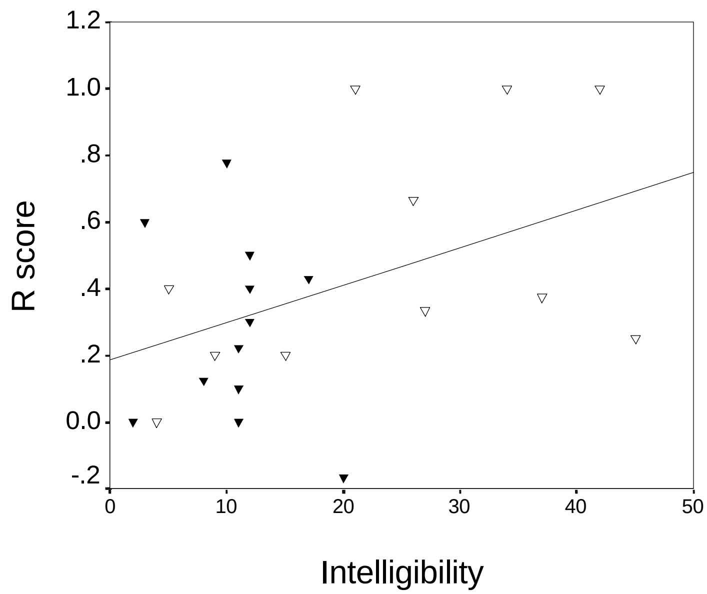

Figure 4. The relationship between integrative ability and intelligibility split by communication mode. OC children are represented by open triangles; TC children are represented by filled triangles.

We also examined the effects of communication mode on R. Eleven of the children who were tested using our measure of speech production were being educated in the Oral Communication (OC) method, while 12 of the children were being educated using the Total Communication (TC) method. Children educated in an oral program are taught to use speaking and listening skills (including lipreading) for communication. Children in this study educated in a TC approach use a combination of spoken and manually-coded English for communication. Figure 4 shows a scatterplot displaying the relationship between intelligibility and visual enhancement (R) for these two groups of children. The diagonal line through the scatterplot is the regression line predicting R from speech intelligibility. For this scatterplot, the data are split by communication mode, with filled triangles indicating children who use TC and open triangles representing children who use OC. As indicated by the positive correlation noted above, higher intelligibility scores were associated with higher R scores. In addition, it appears that most of the high scorers on the intelligibility scale (and consequently the R scale) used OC. In fact, 72.7% of the OC children in this figure were above the median intelligibility score of 12%. In contrast, only 41.7% of the TC children achieved speech intelligibility scores higher than the median.

Discussion

The present study examined the ability of prelingually deafened children with cochlear implants to integrate information about speech from multiple sensory modalities. The results of our analyses demonstrate that these multimodal integration skills are not isolated or independent but are related to more general, unimodal speech perception and production abilities. We observed a strong relationship between R score, a measure of visual enhancement, and the LNT Easy, MLNT Easy, and MLNT Hard tests. Better audiovisual integrators were also better performers on auditory-alone measures of spoken word recognition. In addition, we found a positive correlation between R and speech intelligibility: children who were better AV integrators produced more intelligible speech.

Audiovisual integration reflects the ability of perceivers to combine and use information from multiple sensory modalities to recognize spoken words, syllables, and phonemes (Braida, 1991; Fowler & Dekle, 1991; Green & Gerdeman, 1995; Green & Kuhl, 1991; Kuhl & Meltzoff, 1984; Massaro & Cohen, 1995; Remez et al., 1999; Rosenblum & Saldaña, 1996; Summerfield, 1987; Vatikiotis-Bateson, Munhall, Hirayama, Lee, & Terzepoulos, 1997). Recent theoretical accounts of AV integration assume that auditory and visual information about speech are integrated by the perceptual system because both carry information relevant to the dynamic behavior of articulating vocal tracts (Rosenblum & Saldaña, 1996). Visual access to the action of the lips, the tongue tip and even the jaw can provide useful information about the behavior of the vocal tract. Auditory access to the action of more internal articulators, such as the tongue blade, the tongue body, and the velum provides information about the same articulatory events (Summerfield, 1987). It is precisely this time-varying articulatory behavior of the vocal tract that has been shown to be of primary importance in the perception of speech (Liberman & Mattingly, 1985; Remez, Rubin, Berns, Pardo, & Lang, 1994; Remez, Rubin, Pisoni, & Carrell, 1981).

Viewed with this theoretical context, the information relevant for speech perception is said to be *modality-neutral*, since it can be carried by more than one sensory modality. In addition, because the acoustic and optic specifications of speech information are produced by the same, underlying articulatory events, they are lawfully related to each other and to the underlying events that produce them (Vatikiotis-Bateson et al., 1996). Consequently, as long as information about the articulations of the vocal tract can be perceived, some degree of speech perception will be possible. Indeed, this fact is clearly demonstrated by the remarkable speechreading abilities of some people with hearing impairments (Rönnberg et al., 1999), and even by those with normal hearing (Bernstein, Demorest, & Tucker, in press). Even more impressive is the finding that information obtained via the *tactile* modality can be used and integrated across modalities during speech perception (Fowler & Dekle, 1991), albeit with limited utility.

However, while the *information* necessary for speech perception may be amodal, the internal *representation* of speech must be based on an individual's experience with the sensory world. Clearly, awareness of the intermodal relationships between auditory and visual information must be contingent on experience with more than one sensory modality. Over time, processes of perceptual learning will associate lawful co-occurences in disparate modalities, until a rich, multimodal representation of speech is obtained.

For prelingually deafened children with cochlear implants, however, these perceptual learning processes only begin to develop after they receive their implant and begin to experience the lawful relationships between sights and sounds in their environment. We assume that the degree to which an individual can integrate information across modalities reflects the degree to which they have internalized the inherent, one-to-one relationships between auditory and visual information about speech. The correlations between audiovisual integrative ability and performance on auditory-alone measures of speech perception for children with cochlear implants indicate that the ability to integrate multimodal sources of information reflects a generalized ability to utilize speech information in *any* of its forms, including in unimodal auditory specifications. This conclusion is further supported by the finding that the *productive* capabilities of children with cochlear implants, as measured by speech intelligibility, is significantly correlated to their ability to *perceptually* integrate multimodal sources of information.

We suggest that the generalized ability to utilize articulatory information during speech perception and production arises from rich, more fully specified, multimodal internal representations of the articulatory behavior of the vocal tract. At present, we can only speculate as to how these desirable representations are obtained, and why there might be individual differences in the ability to form them. However, the distribution of TC and OC children in the scatterplot of the relationship between intelligibility and integration ability suggests that early experience that focuses on the vocal communication of language may help to strengthen and solidify speech representations in memory. Teaching hearing-impaired children about language *in general* does not appear to be sufficient for building the kinds of representations most advantageous for vocal speech perception. Rather, a focus on the articulatory events that produce speech, by training orally and aurally, leads to improved ability in the perception and production of spoken language. The relationships observed in the present study between audiovisual integrative ability on the one hand and performance measures of spoken word recognition and speech intelligibility on the other hand suggests that treatment programs should concentrate on teaching children with CIs about the inherent cross-correlations between auditory and visual information in speech. In this way, more robust mental representations of the kinds of events that produce speech will be formed, ultimately leading to more robust performance.

General Discussion and Conclusions

The findings from the present investigation suggest that larger gains in spoken word recognition and language comprehension performance might be obtained if deaf children with CIs were forced to use all sources of sensory information during early stages of perceptual learning after implantation. Our results also suggest that these gains might be more readily acquired if intervention programs, like some oral/aural programs, emphasized the robust multimodal nature of speech events. At the present time, most assessments of performance are often done using “auditory-only” presentations of test materials. For some children who are good lip-readers and who display large gains in visual enhancement, these tests may not provide an accurate assessment of their “true” perceptual skills because an important component and source of sensory information has been arbitrarily removed for purposes of assessment.

The finding of a strong relationship between audiovisual integration performance in perception (as measured by R) and speech intelligibility in some way suggests that the underlying factors that differentiate good CI users from poor CI users must be related to a common underlying set of phonological processing skills. These skills include perceptual, cognitive, and linguistic processes that are involved in the initial encoding, storage, rehearsal and manipulation of phonological representations of spoken words and the construction of sensory-motor programs for speech production and articulation. It is difficult to imagine any theoretical account of the present findings that is framed entirely in terms of peripheral sensory factors related to audibility without the additional assumption of some kind of phonological and/or lexical representation, used in conjunction with some type of linguistic mediation. To account for the present findings, it is necessary to assume the presence of some *underlying linguistic structure and process* which mediates between speech perception and speech production. Without a common underlying linguistic system - a grammar - these separate perceptual and productive abilities would not be so closely coordinated and mutually interdependent. It is well known and clearly documented in the literature on language development that reciprocal links exist between speech perception, production and a whole range of language-related abilities and skills. These links reflect the child's developing linguistic knowledge of phonology, morphology and syntax and his/her attempts to use this knowledge productively in a variety of expressive language tasks.

It needs to be emphasized strongly here that these children have been deprived of auditory experience for a substantial portion of their early lives before they received a cochlear implant. Some of these children are able to quickly learn to perceive and understand spoken language via electrical stimulation from their cochlear implant. Although they are delayed developmentally relative to their normal-hearing peers, these children appear to be making large gains in language development relative to other deaf children who have not received a CI. However, other children with CIs are not so fortunate and they seem to have much more difficulty making use of the sensory information provided by their implant. We believe that the study of individual differences in outcome and effectiveness of CIs is one of the most important research problems to investigate and understand over the next few years. If we had a better understanding of the reasons for these individual differences, we would be in a much better position to recommend changes in the child's language learning environment that were based on some theoretical motivation and set of operating principles. At the present time, decisions are made about intervention and therapy without a firm understanding of exactly why some children do well with their implant and why other children do more poorly.

The findings from the present analysis suggest that the gains in visual enhancement and the excellent audiovisual integration abilities observed in some of these children reflect the development and operation of their underlying linguistic systems. In this connection, several recent investigations have pointed to a close relationship between language and working memory. Pisoni and Geers (1998) reported that differences in working memory may be the "locus" of the wide range of variation observed in children with cochlear implants. Among other findings, they observed a strong positive correlation between auditory digit span and measures of performance on the Chive, a test of audiovisual integration that is similar to the Common Phrases Test used in the present analysis. In an analysis of data collected from a group of 43 eight- and nine-year old prelingually deafened children who had used their CI at least five years, Pisoni and Geers (1998) found that the WISC forward digit span was correlated with Chive V, a measure of lip-reading ability ($r = +0.52$) and Chive VE, a measure of visual enhancement abilities ($r = +0.66$). These findings were interpreted as support for the proposal that the large individual differences observed among children with CIs may be related to fundamental information processing operations in working memory, specifically, the operation of phonological working memory and the "phonological loop" which has been hypothesized as the primary rehearsal mechanism used to code and maintain the phonological representations of spoken words (Baddeley, Gathercole, & Papagno, 1998; Gathercole, Hitch, Service, & Martin, 1997).

Pisoni and Geers' (1998) findings on the role of working memory suggest that an important source of variance in outcome performance in these children has to do with the processing operations in working memory that a child uses to encode the sensory information he/she receives through the cochlear implant. The presence of a correlation between WISC auditory digit spans and visual enhancement due to lip-reading suggests that the locus of these effects may reside in a common representation format that is independent of the specific input modality. In some sense, it doesn't matter if the sensory information is visual or auditory. What matters to the perceiver is whether the sensory input helps to specify the source of the underlying perceptual event, facilitating the recovery of the talker's articulatory gestures.

The results from Pisoni and Geers (1998) on auditory digit span and the present findings on AV integration are consistent with current theoretical conceptualizations of memory. Many theorists view short-term memory as simply that portion of long-term memory that is currently active (Anderson & Bower, 1973; Atkinson & Shiffrin, 1971; Engle, 1996) and closely related to attentional processes. As a consequence, long-term memory representations of spoken words and lexical knowledge play an important role in tasks that involve working memory. We suggest that the relationship observed in the current study between audiovisual integration and auditory-alone measures of speech perception reflects the contribution of fully specified, multimodal articulatory representations of speech in long-term memory and knowledge of the cross-correlations between auditory and visual information about the same underlying articulatory events.

We speculate that the relationship between representational specificity and memory span arises because working memory plays a role in the formation of rich, multimodal representations in memory. Attention has been shown to play a large role in the implicit learning of artificial grammars (Nissen & Bullemer, 1987) and new verbal associations (Hartman, Knopman, & Nissen, 1989). Indeed, it is thought that the underlying statistical structure of patterns in the world facilitates the learning of those patterns (Reber, 1993; Stadler, 1993). If working memory is involved in the acquisition and internalization of statistical form in the world, then the relationships we have observed in this and other studies have a coherent explanation. Becoming a good integrator very likely entails perceptual learning and the internalization and representation of natural, lawful co-occurrences between disparate sensory information channels. Further work on this issue is needed and is currently being planned for our lab.

In summary, the present study uncovered relationships between the ability of children with cochlear implants to integrate cross-modal information about speech and their performance on auditory-alone measures of spoken word recognition. In addition, we found a close relationship between a child's audiovisual integrative performance and speech intelligibility. We conclude that these relationships point to the importance of building rich, multimodal long-term memory representations of speech that emphasize the amodal nature of speech information. Because these representations are necessary for a variety of speech tasks, performance on those tasks improves with complementary sensory inputs. We suspect that the most fruitful treatment programs for the speech perception and production of children with cochlear implants will involve emphasis on the vocal source of spoken language.

References

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, DC: Winston.

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American, 225*, 82 - 90.

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158 - 173.

Bernstein, L. E., Demorest, M. E., & Tucker, P. E. (in press). Speech perception without hearing. *Perception & Psychophysics*.

Bollard, P. M., Chute, P. M., Popp, A., & Parisier, S. C. (1999). Specific language growth in young children using the Clarion cochlear implant. *Annals of Otology, Rhinology, & Laryngology, 108*(Suppl. 117), 119 - 123.

Braida, L. D. (1991). Crossmodal integration in the identification of consonant segments. *Quarterly Journal of Experimental Psychology, 43A*(3), 647 - 677.

Dowell, R. C., Martin, L. F. A., Tong, Y. C., Clark, G. M., Seligman, P. M., & Patrick, J. E. (1982). A 12-consonant confusion study on a multiple-channel cochlear implant patient. *Journal of Speech & Hearing Research, 25*, 509 - 516.

Engle, R. W. (1996). Working memory and retrieval: An inhibition-resource approach. In J. T. E. Richardson (Ed.), *Working memory and human cognition* (pp. 89 - 119). Oxford: Oxford University Press.

Erber, N. P. (1972). Auditory, visual and auditory-visual recognition of consonants by children with normal and impaired hearing. *Journal of Speech, Language, and Hearing Research, 15*, 413 - 422.

Fowler, C. A. (1989). Real objects of speech perception: A commentary on Diehl and Kluender. *Ecological Psychology, 1*(2), 145 - 160.

Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception & Performance, 17*(3), 816 - 828.

Fryauf-Bertschy, H., Tyler, R. S., Kelsay, D. M., & Gantz, B. J. (1992). Performance over time of congenitally deaf and postlingually deafened children using a multichannel cochlear implant. *Journal of Speech & Hearing Research, 35*, 913 - 920.

Fryauf-Bertschy, H., Tyler, R. S., Kelsay, D. M. R., Gantz, B. J., & Woodworth, G. G. (1997). Cochlear implant use by prelingually deafened children: The influences of age at implant and length of device use. *Journal of Speech & Hearing Research, 40*, 183 - 199.

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33*, 966 - 979.

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *Journal of the Acoustical Society of America, 103*(5), 2677 - 2690.

Green, K. P., & Gerdeman, A. (1995). Cross-modal discrepancies in coarticulation and the integration of speech information: The McGurk effect with mismatched vowels. *Journal of Experimental Psychology: Human Perception and Performance, 21*(6), 1409 -1426.

Green, K. P., & Kuhl, P. K. (1991). Integral processing of visual place and auditory voicing information during phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 17*(1), 278 - 288.

Hartman, M., Knopman, D. S., & Nissen, M. J. (1989). Implicit learning of new verbal associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(6), 1070 - 1082.

Haskins, H. (1949). *A phonetically balanced test of speech discrimination for children.* Unpublished master's thesis, Northwestern University, Evanston, IL.

Kirk, K. I. (1999). Assessing speech perception in listeners with conchlear implants: The development of the lexical neighborhood tests. *The Volta Review, 100*(2), 63 - 85.

Kirk, K. I., Diefendorf, A. O., Pisoni, D. B., & Robbins, A. M. (1997). Assessing speech perception in children. In L. L. Mendel & J. L. Danhauer (Eds.), *Audiologic evaluation and management and speech perception assessment* (pp. 101 - 132). San Diego: Singular Publishing.

Kirk, K. I., Eisenberger, L. S., Martinez, A. S., & Hay-McCutcheon, M. (1999). The Lexical Neighborhood Tests: Test-retest reliability and interlist equivalency. *Journal of the American Academy of Audiology, 10*, 113 - 123.

Kirk, K. I., Pisoni, D. B., & Osberger, M. J. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear & Hearing*, *16*, 470 - 481.

Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant behavior and development, 7,* 361 - 381.

Liberman, A., & Mattingly, I. (1985). The motor theory revised. *Cognition, 21*, 1 - 36.

Logan, J. S. (1992). *A computational analysis of young children's lexicons* (Research on Speech Perception Technical Report No. 8). Bloomington, IN: Speech Research Laboratory, Indiana University.

Luce, P. A. (1986). Neighborhoods of words in the mental lexicon, *Research on Speech Perception Technical Report No. 6* . Bloomington, IN: Speech Research Laboratory, Indiana University.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear & Hearing, 19,* 1 - 36.

Massaro, D. W., & Cohen, M. M. (1995). Perceiving talking faces. *Current Directions in Psychological Science*, *4*(4), 104-109.

Massaro, D. W., & Cohen, M. M. (1999). Speech perception in perceivers with hearing loss: Synergy of multiple modalities. *Journal of Speech, Language, and Hearing Research, 42*, 21 - 41.

Meyer, T. A., & Pisoni, D. B. (1999). Some computational analyses of the PBK test: Effects of frequency and lexical density on spoken word recognition. *Ear & Hearing, 20*, 363 - 371.

Miyamoto, R. T., Osberger, M. J., Robbins, A. M., Renshaw, J. J., Myres, W. A., Kessler, K., & Pope, M. L. (1989). Comparison of sensory aids in deaf children. *Annals of Otology, Rhinology, & Laryngology, 98,* 2 - 7.

Miyamoto, R. T., Svirsky, M. A., Kirk, K. I., Robbins, A. M., Todd, S., & Riley, A. (1997). Speech intelligibility of children with multichannel cochlear implants. *Annals of Otology, Rhinology, & Laryngology, 106,* 35 - 36.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19,* 1 - 32.

Osberger, M. J., Miyamoto, R. T., Zimmerman-Phillips, S., Kernink, J. K., Stroer, B. S., Firzst, J. B., & Novak, M. A. (1991). Independent evaluation of the speech perception abilities of children with the Nucleus 22-channel cochlear implant system. *Ear & Hearing, 12*(Suppl.), 66S -80S.

Pisoni, D. B., & Geers, A. (1998). Working memory in deaf children with cochlear implants: Correlations between digit span and measures of spoken language processing, *Research on Spoken Language Processing Progress Report No. 22* (pp. 335 - 343). Blooomington, IN: Speech Research Laboratory, Indiana University.

Pisoni, D. B., Svirsky, M. A., Kirk, K. I., & Miyamoto, R. T. (submitted). Looking at the "stars": A first report on the intercorrelations among measures of speech perception, intelligibility and language in pediatric cochlear implant users. *Journal of Speech, Language, and Hearing Research*.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious.* New York, NY: Oxford University Press.

Remez, R. E., Fellowes, J. M., Pisoni, D. B., Goh, W. D., & Rubin, P. E. (1999). Multimodal perceptual organization of speech: Evidence from tone analogs of spoken utterances. *Speech Communication, 26*(1-2), 65 - 73.

Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review, 101*(1), 129-156.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212*, 947 - 950.

Robbins, A. M., Bollard, P. M., & Green, J. (1999). Language development in children implanted with the Clarion cochlear implant. *Annals of Otology, Rhinology, & Laryngology, 108*(Suppl. 117), 113 - 118.

Robbins, A. M., Svirsky, M. A., & Kirk, K. I. (1997). Children with implants can speak, but can they communicate? *Otolaryngology-Head and Neck Surgery, 117,* 155 - 160.

Rönnberg, J., Andersson, J., Samuelsson, S., Söderfeldt, B., Lyxell, B., & Risberg, J. (1999). A speechreading expert: The case of MM. *Journal of Speech, Language, and Hearing Research, 42*, 5 - 20.

Rosenblum, L. D. (1994). How special is audiovisual speech integration? *Current Psychology of Cognition, 13*(1), 110 - 116.

Rosenblum, L. D., Johnson, J. A., & Saldaña, H. M. (1996). Point-light facial displays enhance comprehension of speech in noise. *Journal of Speech, Language, and Hearing Research, 39*, 1159 - 1170.

Rosenblum, L. D., & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception & Performance*, *22*(2), 318 - 331.

Stadler, M. A. (1993). Implicit serial learning: Questions inspired by Hebb (1961). *Memory and Cognition, 21*(6), 819 - 827.

Staller, S., Beiter, A. L., Brimacombe, J. A., Mecklenberg, D., & Arndt, P. (1991). Pediatric performance with the Nucleus 22-channel implant system. *American Journal of Otology, 12*, 126 - 136.

Sumby, W. H., & Pollack, I. (1954). Visual contribution of speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212 - 215.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by Eye: The Psychology of Lip-Reading* (pp. 3 - 51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Svirsky, M. A., Sloan, R. B., Caldwell, M., & Miyamoto, R. T. (1998). *Speech intelligibility of prelingually deaf children with multichannel cochlear implants.* Paper presented at the Seventh Symposuim on Cochlear Implants in Children, Iowa City, IA.

Tyler, R. S., Fryauf-Bertschy, H., Gantz, B. J., Kelsay, D. M. R., Tyler, R. S., Woodworth, G. G., & Parkinson, A. (1997a). Speech perception by prelingually deaf children using cochlear implants. *Otolaryngology-Head and Neck Surgery, 117,* 180 -187.

Tyler, R. S., Fryauf-Bertschy, H., Gantz, B. J., Kelsay, D. M. R., & Woodworth, G. G. (1997b). Speech perception in prelingually implanted children after four years. In I. Honjo & H. Takahashi (Eds.), *Cochlear implant and related science update (Advances in Otolaryngology 52)* (pp. 187 - 192). Basel: Karger.

Tyler, R. S., Parkinson, A. J., Fryauf-Bertschy, H., Lowder, M. W., Parkinson, W. S., Gantz, B. J., & Kelsay, D. M. R. (1997c). Speech perception by prelingually deaf children and postlingually deaf adults with cochlear implants. *Scandinavian Journal of Audiology, 26*(Suppl.46), 65 - 71.

Tyler, R. S., Tomblin, J. B., Spencer, L. J., Kelsay, D. M. R., & Fryauf-Bertschy, H. (in press). How speech perception through a cochlear implant affects language and education. *Otolaryngology-Head and Neck Surgery*.

Vatikiotis-Bateson, E., Munhall, K. G., Hirayama, M., Lee, Y. V., & Terzepoulos, D. (1997). The dynamics of audiovisual behavior in speech. In D. G. Stork & M. E. Hennecke (Eds.), *Speechreading by Humans and Machines* (pp. 221 - 232). Berlin: Springer-Verlag.

Vatikiotis-Bateson, E., Munhall, K. G., Kasahara, Y., Garcia, F., & Yehia, H. (1996). *Characterizing audiovisual information during speech.* Paper presented at the International Conference on Spoken Language Processing, Philadelphia, PA.

Zimmerman-Phillips, S., Osberger, M. J., & Robbins, A. M. (1997). *Infant toddler meaningful auditory integration scale (IT-MAIS)*. Sylmar, CA: Advanced Bionics Corporation.