Word Recognition by Children Listening to Speech Processed into a Small Number of Channels: Data from Normal-Hearing Children and Children with Cochlear Implants

Michael F. Dorman, Philipos C. Loizou, Lauren L. Kemp, and Karen Iler Kirk

Objective: **The aims of this study were 1) to determine the number of channels of stimulation needed by normal-hearing adults and children to achieve a high level of word recognition and 2) to compare the performance of normal-hearing children and adults listening to speech processed into 6 to 20 channels of stimulation with the performance of children who use the Nucleus 22 cochlear implant.**

Design: **In Experiment 1, the words from the Multisyllabic Lexical Neighborhood Test (MLNT) were processed into 6 to 20 channels and output as the sum of sine waves at the center frequency of the analysis bands. The signals were presented to normal-hearing adults and children for identification. In Experiment 2, the wideband recordings of the MLNT words were presented to early-implanted and late-implanted children who used the Nucleus 22 cochlear implant.**

Results: **Experiment 1: Normal-hearing children needed more channels of stimulation than adults to recognize words. Ten channels allowed 99% correct word recognition for adults; 12 channels allowed 92% correct word recognition for children. Experiment 2: The average level of intelligibility for both early- and late-implanted children was equivalent to that found for normal-hearing adults listening to four to six channels of stimulation. The best intelligibility for implanted children was equivalent to that found for normal-hearing adults listening to six channels of stimulation. The distribution of scores for early- and late-implanted children differed. Nineteen percent of the late-implanted children achieved scores below that allowed by a 6-channel processor. None of the early-implanted children fell into this category.**

Conclusions: **The average implanted child must deal with a signal that is significantly degraded. This is likely to prolong the period of language acquisition. The period could be significantly shortened if implants were able to deliver at least eight functional channels of stimulation. Twelve functional channels of stimulation would provide signals near the intelligibility of wideband signals in quiet.**

(Ear & Hearing 2000;21;590–596)

One of the central questions in the design of a cochlear implant is how many channels of stimulation are needed to achieve a high level of speech understanding. One way to gain a purchase on the answer to this question is to first answer a different question, i.e., how many channels of stimulation are needed by normal-hearing listeners to achieve a high level of speech understanding. The answer sets a minimum criterion for the number of channels that should be provided to an implant patient.

Recently two groups of investigators have assessed the number of channels needed by normal-hearing adults to achieve a high level of speech understanding. Both sets of investigators used similar signal processing techniques that were modeled after the signal processing algorithms for a cochlear implant. Signals were divided into N bands, the energy in each band was estimated over short time intervals, and either noise bands the width of the analysis filters or sine waves at the center frequency of each analysis band were output with amplitudes proportional to the energy in the bands. The results for noise band outputs and sine wave outputs were similar. To reach 90% accuracy in quiet, four channels are necessary for simple sentences (Dorman, Loizou, & Rainey 1997; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Five channels are necessary for more difficult sentences produced by multiple talkers (Loizou, Dorman, & Tu, 1999) and eight or more channels are necessary for monosyllablic words (Dorman, Loizou, Fitzke, & Tu, in press).

In the research reported here we extend our observations on the number of channels necessary to reach high levels of word understanding to young (3- to 5-yr-old), normal-hearing children. Young children do not have the full complement of linguistic and cognitive skills enjoyed by an adult. As a consequence young children may differ from adults in how

Arizona State University (M.F.D., L.L.K.), Tempe, Arizona; University of Utah Health Sciences Center (M.F.D.), Salt Lake City, Utah; University of Texas at Dallas (P.C.L.), Dallas, Texas; and Indiana University School of Medicine (K.I.K.), Indianapolis, Indiana.

they respond to signals that have reduced spectral content.

When assessing the effects of reduced spectral content on word recognition it is essential, of course, to use words that are in the vocabulary of the children. A significant number of the words that occur in one of the most common tests for children, the Phonetically Balanced Kindergarten word lists (Haskins, Reference Note 1), are not high-frequency words in the vocabularies of present-day children (Kirk, 1999; Kirk, Pisoni, & Osberger, 1995). To remedy this problem, Kirk et al. (1995) developed tests of word recognition for children based on recent analyses of a database of verbal exchanges between children and a caretaker or other children, the Child Language Data Exchange System or CHILDES (MacWhinney & Snow, 1985). For each test (the Lexical Neighborhood Test and the Multisyllabic Lexical Neighborhood Test [MLNT]), two word lists, one containing "hard" words and one containing "easy" words, were developed. The "easy" words were selected to be above the median for word frequency and below the median for "neighborhood density" for words in the CHILDES database whereas the "hard" words had the opposite characteristics. Neighborhood density refers to the number of words that could be found in the CHILDES database by adding, substituting or deleting one phoneme from the target word. Thus, "easy" words have few phonetically similar neighbors. "Hard" words have many neighbors.

EXPERIMENT 1

In Experiment 1 the "easy" and "hard" words from the MLNT were used to assess whether young children were as capable as adults of understanding speech that had been processed into a small number of channels.

Method

Subjects • The subjects were 36 children between the ages of 3.5 yr and 5.7 yr and 36 young adults. The children had normal hearing and normal language development by parental and teacher report. The children were recruited from the on-campus preschools at Arizona State University. The young adults were undergraduate and graduate students at Arizona State University.

Test Materials and Signal Processing • The "easy" and "hard" word lists from the MLNT were used as test material. There were 24 "easy" and 24 "hard" words in the test set.

To create 4-, 6-, 8-, 10-, and 12-channel test materials, the words from the MLNT were processed in the following manner. The signals were first processed through a preemphasis filter (low-pass below 1200 Hz with 6 dB/octave rolloff) and then band passed into N frequency bands (where N varied from 6 to 12) using sixth-order Butterworth filters. The spacing of the frequency bands was logarithmic for the 4-, 6-, and 8-channel processors. For the 10- and 12-channel processors, linear spacing of bands was used up to 1 kHz and logarithmic spacing thereafter. The envelope of the signal was extracted by full-wave rectification, and low-pass filtering (second-order Butterworth) with a 400 Hz cutoff frequency. Sinusoids were generated with amplitudes equal to the root-mean-square (rms) energy of the envelopes (computed every 4 msec) and frequencies equal to the center frequencies of the band-pass filters. The sinusoids were summed and presented to the listeners at a comfortable level (72 dB SPL re: vowel peaks).

One additional set of test material was created. For this set, signals were processed in the manner of a "channel-picking" speech processor, one similar to the SPEAK algorithm used in the Nucleus cochlear implant (McDermott, McKay, & Vandali, 1992). Signals were processed into 20 channels, using linear spacing to 1 kHz and logarithmic spacing thereafter, and every 4 msec the six channels with the maximum energy were identified and output as sine waves at the center frequency of the analysis band.

For each channel condition, the 24 "hard" words and 24 "easy" words were interleaved into a test sequence. Interleaving assured equal representation of "easy" and "hard" words in the beginning, middle and end of the test list. The stimuli were presented monaurally using custom software from a laptop computer with 16 bit D/A converters via Sennheiser HMD410 headphones.

Procedures • Six adults and six children were randomly assigned to each of the six channel conditions (i.e., 4, 6, 8, 10, 12, and 20 channels). The number of relatively younger children and relatively older children was balanced across groups. The subjects were tested in a quiet room (not in a sound-attenuated chamber). The subjects were told that they would hear words over headphones. They were told to write (adults) or say (children) what was heard and to guess if necessary. Practice consisted of listening to 10 sentences from the HINT lists (Nilsson, Soli, & Sullivan, 1994) and 20 words from the Lexical Neighborhood Test processed into 12 channels. The practice material served also as a screening test. All of the children were able to repeat the sentences and 75% of the words in the practice list. The youngest children (age 3.5 yr) performed as well as the older children (5.7 yr) on this task.

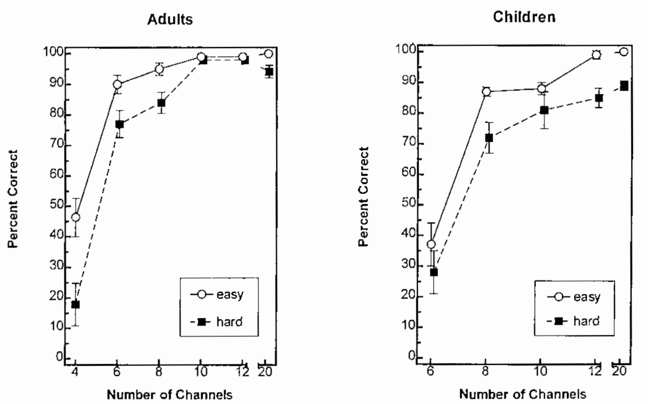

Figure 1. Percent correct word recognition as a function of the number of channels of fixed-frequency, sine wave stimulation for adults (left panel) and for children (right panel). Error bars indicate ±1 standard error of the mean.

Results

The performance of the adults is shown in Figure 1 on the left and in Figure 2. There was a main effect for number of channels ($F[5,71] = 67.36$, $p < 0.00001$) and for lexical difficulty ($F[1,71] = 78.01$, $p < 0.00001$). There was a significant interaction between the two main effects ($p < 0.000001$). Maximum performance for the "easy" and "hard" words (99% and 98% correct, respectively) was reached with 10 channels of stimulation.

The performance of the children is shown in Figure 1 on the right and in Figure 2. The 4-channel condition was so difficult that none of the children finished the task. For that reason the 4-channel condition was dropped from the data analyses. There was a main effect for number of channels ($F[4,59] = 51.07$, $p < 0.00001$) and for lexical difficulty ($F[1,59] = 32.54$, $p < 0.00006$). The interaction term was not significant ($p = 0.08$). Maximum performance for the "easy" words (99 to 100% correct) was reached with the 12-channel processor and the 6 of 20 processor. Maximum performance for the "hard" words (89% correct) was achieved with the 6 of 20 processor.

Discussion

Overall Performance • The results shown in Figure 2 indicate that, most generally, children perform at a lower level than adults when presented with signals that have been processed into a small number of channels. As a consequence, children generally need more channels than adults to reach a given level of performance. As shown in Figure 2, children need about four more channels than adults to reach 80 to 90% correct. The deficit in performance for the children may be attributed, on the one hand, to factors that are narrowly related to lexical access or, on the other hand, to nonlinguistic, task-related factors. The latter does not provide a general account of performance because the children's performance was equal to that of the adults for the "easy" words processed into 12 and 20 channels*. Thus, it is likely that the differences in performance between the children and adults are due, principally, to differences in achieving lexical access from signals with reduced spectral information. The differences in ability to achieve lexical access from signals with reduced information could be due to 1) a less robust pattern matching algorithm for children or 2) a more narrow representation, or template, of a word in a child's memory (see Lively, Pisoni, & Goldinger, 1994, for a review of theories of spoken word recognition).

The finding that children are less proficient than adults in recognizing words from degraded sensory information is not unique. Our results are similar in kind to results from experiments with children in which sensory information is degraded by noise (e.g., Elliot, 1979), by reverberation (e.g., Neuman & Hochberg, 1983) and by time compression (e.g., Nagafuchi, 1976). Moreover, Eisenberg, Martinez, Wygonski, and Boothroyd (2000) have conducted an experiment similar in design to the one reported here and have found similar results. Vowel-consonant-vowel segments, words and sentences were processed into 4 to 32 channels and output as bands of noise the width of the analysis channels to children aged 5 to 7 yr, to children aged 10 to 12 yr and to young adults. The youngest children needed more channels than the older children and adults to reach the same level of performance. Additional analyses suggested that the youngest children were not able to use sentence context as well as the older children and adults in the pursuit of word recognition. Thus, children need more information (channels) at the sensory level for lexical access and are less able to use sentence context for lexical access.

Easy versus Hard Words • For easy words, the children matched the level of performance of the adults at 12 channels of stimulation (although performance was constrained by a ceiling effect). For hard words, the children did not reach the level of performance of the adults even with the 6 of 20 processor. These findings can be interpreted in the following way. Even if pattern matching algorithms

*The children had listened to 12-channel speech during the practice session and, as a consequence, had more practice with type of signal than the other types of signals. It is possible, though very unlikely, that the asymptote in performance at 12 channels was due to the extra familiarity with speech processed into 12 channels. We note that performance with the 6 of 20 processor was also at asymptote, and the children did not have extra practice with the signals produced by this processor.

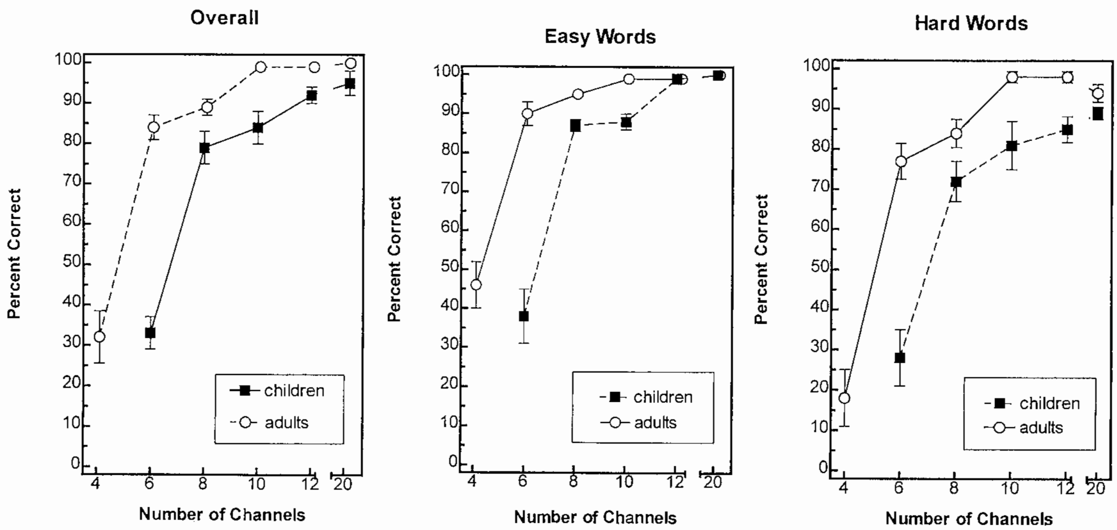

Figure 2. Percent correct word recognition as a function of the number of channels of stimulation. Left panel: performance averaged over "easy" and "hard" words. Center panel: performance on "easy" words. Right panel: performance on "hard" words.

are less robust (or templates are more specific) in children, the absence of lexical neighbors for the easy words allows recognition to take place with 12 channels. On the other hand, the presence of many neighbors for the hard words ensures that a better representation of the signal is needed to generate the correct match. Children need a very detailed representation of the input signal (more than 20 input channels and six output channels) to achieve scores of 95 to 100% for words in a densely populated neighborhood.

Degraded information and language acquisition • The normal-hearing children in the present experiment and in Eisenberg et al. (2000) found signals processed into six channels or less difficult to identify. Several lines of research with adults suggest that adult implant patients obtain the equivalent of only four to six channels of information from a speech signal even if they have as many as 22 implanted electrodes (Dorman et al., in press; Fishman, Shannon, & Slattery, 1997; Wilson, 1997). If implanted children, like adults, only extract the equivalent of relatively few channels of stimulation from signals processed into a much larger number of channels, the cortical pattern recognition processes in implanted children must form representations of words, syllables or phonetic units based on relatively degraded information. It is likely that this circumstance would prolong, or make more difficult, the task of language acquisition.

The task of language acquisition would be facilitated if an implant could provide children with signals that were equivalent, or nearly equivalent, in terms of intelligibility to wideband speech. The results of Experiment 1 suggest that 12 channels of stimulation allow a level of performance in quiet, 92% correct overall, near that of wideband speech.

If 12 channels cannot be delivered to an implanted child, stimulation through eight channels is a reasonable goal. Performance with eight channels, while not at asymptote, is good especially for "easy" words. If we could provide the equivalent of eight channels of sine wave stimulation, the task of lexical access might be greatly facilitated and the rate of language acquisition might be increased. Indeed, it is probably the case that the best performing children on tests of language acquisition are children who receive the equivalent of at least eight sine wave channels of stimulation (Svirksy, Robbins, Kirk, Pisoni, & Miyamoto, 2000).

EXPERIMENT 2

The aim of Experiment 2 was to compare the word recognition performance of implanted children with the word recognition performance of the normal-hearing children and adults tested in Experiment 1. At issue was whether implanted children, even the best performing children, receive all, or most, of the information potentially available from an implant's signal processor.

The design of this experiment was similar to one used previously in a study with adults (Dorman et al., in press). In that study, monosyllabic word scores for adults who used the 8-channel Clarion device, the Nucleus 22 and 24 devices and the 8-channel Med El device were plotted against the scores of normal-hearing adults who listened to the NU6 words processed into 4 to 20 channels of sine wave stimulation. The mean level of performance for the implanted adults was equivalent to that of normal-hearing adults listening to speech processed into four channels. The best performing adults with implants achieved scores in the range of scores for normal-hearing adults listening to speech processed into eight channels. Thus, the average adult implant patient does not receive all, or even most, of the information potentially available from his/her device. Only a small number of patients, approximately 12%, receive most of the information available from their device. At issue in Experiment 2 was whether children would behave like adults and, on average, receive only a limited amount of the information potentially available from signal processors with many channels of stimulation.

Method

Subjects • The subjects were 56 children with Nucleus 22 cochlear implants from the sample at the Indiana University School of Medicine. Forty-three of the children were classified as "late implanted" and 13 of the children were classified as "early implanted."

For the late-implanted children, the mean age at device fitting was 5.4 yr. The mean age at onset of deafness was 0.2 yr. Eighty-two percent of the children were deaf at birth. The mean length of device use was 4.5 yr. The mean number of electrodes activated for the children using the Nucleus device was 17.

For the early-implanted children the mean age at device fitting was 3.3 yr. Thirteen of the 14 children were deaf at birth. One was deafened at 9 mo. The mean length of device use was 3.1 yr.

Test Material • The stimuli were the original, wideband recordings of the words from the MLNT.

Results and Discussion

The results for the normal-hearing children and adults listening to the MLNT words processed into channels (from Experiment 1) and the results for the two groups of implanted children are shown in Figure 3. The mean score for the late-implanted children was 45% correct. The mean score for the early-implanted children was 51%. These levels of performance fall between the levels of performance allowed by 6- and 8-channel processors for normal-hearing children and are closer to the level of performance allowed by a 6-channel processor than an 8-channel processor. One child performed at a level greater than that allowed by an 8-channel processor. Overall, these results indicate that, on average, implanted children, like implanted adults receive only a limited amount of the information potentially available from signal processors with 20 input channels. The results also suggest that children receive at least as much information as adults. Thus, although children are subject to the same factors as adults that limit the use of a large number of channels of stimulation, e.g., channel and current interactions, children do not appear to be more affected by these conditions than adults.

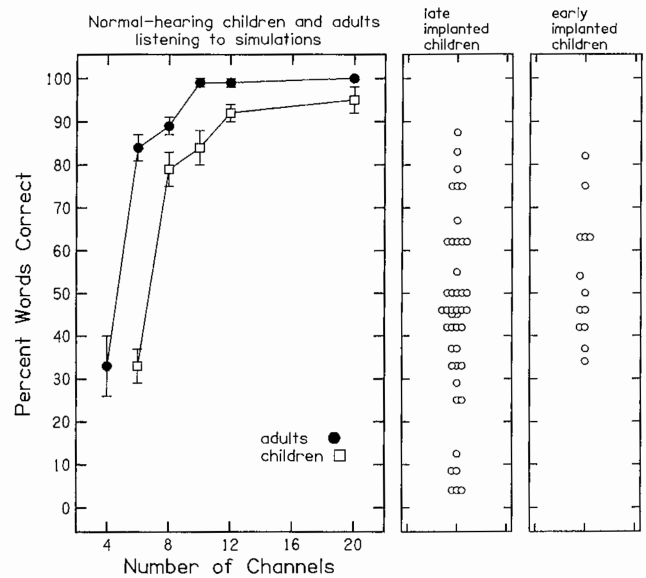

Figure 3. Percent correct word recognition as a function of the number of channels of stimulation for normal-hearing children and adults listening to simulations (left panel). The 20-channel processor was implemented as a 6 of 20 processor. Percent correct word recognition for children using the Nucleus cochlear implant (center and right panel).

Early versus Late • As shown in Figure 3, 19% of the late-implanted children performed at a level below that of a 6-channel processor. Strikingly, none of the early-implanted children performed below the level of a 6-channel processor. To provide converging evidence on this difference between early- and late-implanted children, data from an additional 10 early-implanted children were analyzed. These children had been tested with a multi-talker version of the MLNT instead of the single talker version used in the present experiment. For that reason the patients could not be included in the main analysis of results. Nine of the 10 subjects had scores between 37 and 67% correct with a mean score of 50% correct. One of the 10 subjects scored 0% correct. Thus, the data from this subset of patients are consistent with the data from the larger set of early implanted patients. Overall, the data demonstrate that the largest difference between early and late-implanted children is not the number of very high scores achieved by early implanted patients, but rather the relative absence of very low scores for the early implanted patients. The absence of low scores for the early-implanted children may be linked to better neural survival in this group relative to the later implanted children. On this view, neural survival in some late-implanted children is so poor that the children receive the equivalent of only a few channels of stimulation from their processor.

Adults and Children • Comparisons between implanted children and normal-hearing children are useful, but do not allow an adult to gain an impression of the intelligibility of speech achieved by a child with an implant. To achieve this end we should compare the data from implanted children with the data from adults. The mean level of performance for the implanted children falls between the levels of performance allowed by 4- and 6-channel processors for adults and is closer to four than six. The scores for all but one of the best performing children falls within the standard deviation of scores for adults listening to speech through a 6-channel processor.

One virtue of comparing the performance of the implanted children against that of normal-hearing adults is that an adult can listen to the processed signals and gain an appreciation of the intelligibility (but, importantly, not the subjective quality) of speech allowed by cochlear implants in children. A subset of the MLNT words processed into 4 to 12 channels can be downloaded for listening from www.utdallas.edu/loizou/cimplants/children.

Channels of Stimulation • Consider the outcome that the average score of an implanted child falls between the scores for normal-hearing adults listening to four and six channels. This does not mean that the average implanted child receives useful information from only four or six of the channels of a 22-electrode array. It is possible that children receive fragmentary information from many channels. Yet, it is the case that the sum of the information received produces a percept that could be provided to an adult by a processor with only four or six channels or provided to children by a processor with six or eight channels.

CONCLUSIONS

• Children need more channels than adults to reach a given level of performance on tests of word recognition, e.g., children need four more channels than adults to achieve 80 to 90% correct on the MLNT. The differences in performance between adults and children are most likely due to differences in the ability to access lexical information from speech with degraded spectral cues.

• The mean score for late-implanted children on the MLNT was 45% correct. The mean score for early-implanted children on the MLNT was 51% correct. These levels of performance fall between the levels of performance achieved by normal-hearing children listening to words processed through six and eight channels and falls between the levels of performance achieved by normal-hearing adults listening to words processed through four and six channels.

• The distribution of MLNT scores for the early and late-implanted children differ. Nineteen percent of the late-implanted children achieved scores below that allowed by a 6-channel processor. None of the early-implanted children fell into this category. The relative absence of very poor scores for the early-implanted children suggests a higher probability of neural survival for this group of children. In contrast, the late-implanted children may be characterized by variable neural survival.

• Children with implants, like adults with implants, receive only a portion of the information potentially available from a signal processor with 20 channels. If signal processors/electrode arrays could provide implanted children with the equivalent of 8 to 12 functional channels of stimulation, the tasks of word recognition and language acquisition would be significantly facilitated.

ACKNOWLEDGMENTS:

This research was supported by a grant from the National Institute on Deafness and Other Communication Disorders (RO1 DC 000654–8) to M.F.D. and by grants from the National Institute on Deafness and Other Communication Disorders (RO1 DC00064–11 and K23 DC000126–04) and from the Psi Iota Xi National Sorority to KIK. We are indebted to Mario A. Svirsky of the Indiana University School of Medicine for providing the original MLNT speech files.

Address for correspondence: M. F. Dorman, Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102.

Received December 6, 1999; accepted May 19, 2000

REFERENCES

Dorman, M., Loizou, P. Fitzke, J., & Tu, Z. (in press). The recognition of monosyllabic words by cochlear-implant patients and by normal-hearing subjects listening to NU-6 words processed in the manner of CIS and SPEAK strategies. *Annals of Otology, Rhinology and Laryngology*.

Dorman, M. F., Loizou, P., & Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *Journal of the Acoustical Society of America*, *102*, 2403–2411.

Eisenberg, L., Shannon, R., Martinez, A., Wygonski, J., & Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *Journal of the Acoustical Society of America*, *107*, 2704–2710.

Elliot, L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability. *Journal of the Acoustical Society of America*, *66*, 651–653.

Fishman, K., Shannon, R., & Slattery, W. (1997). Speech recognition as a function of the number of electrodes used in the

SPEAK cochlear implant speech processor. *Journal of Speech and Hearing Research*, *40*, 1201–1215.

Kirk, K. (1999). Assessing speech perception in listeners with cochlear implants: The development of the Lexical Neighborhood Tests. *Volta Review*, *100*, 63–85.

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

Lively, S., Pisoni, D., & Goldinger, S. (1994). Spoken word recognition: Research and theory. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 265–301). New York: Academic Press.

Loizou, P., Dorman, M., & Tu, Z. (1999). On the number of channels needed to understand speech. *Journal of the Acoustical Society of America*, *106*, 2097–2103.

MacWhinney, B., & Snow, C. (1985). The child language data exchange system. *Journal of Child Language*, *12*, 271–296.

McDermott, H., McKay, C., & Vandali, A. (1992). A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant. *Journal of the Acoustical Society of America*, *91*, 3367–3371.

Nagafuchi, M. (1976). Intelligibility of distorted speech sounds shifted in frequency and time in normal children. *Audiology*, *15*, 326–337.

Neuman, A., & Hochberg, I. (1983). Children's perception of speech in reverberation. *Journal of the Acoustical Society of America*, *73*, 2145–2149.

Nilsson, M., Soli, S., & Sullivan, J. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and noise. *Journal of the Acoustical Society of America,*, *95*, 1085–1099.

Shannon, R., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

Svirsky, M., Robbins, A., Kirk, K. I., Pisoni, D., & Miyamoto, R. (2000). Language development in profoundly deaf children with cochlear implants. *Psychological Science*.

Wilson, B. (1997). The future of cochlear implants. *British Journal of Audiology*, *31*, 205–225.

REFERENCE NOTES

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