

The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels

Michael F. Dorman

*Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102
and University of Utah Health Sciences Center, Salt Lake City, Utah 84132*

Philipos C. Loizou

Department of Applied Sciences, University of Arkansas at Little Rock, Little Rock, Arkansas 72204-1099

Jeanette Fitzke

Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102

Zhemín Tu

Department of Applied Sciences, University of Arkansas at Little Rock, Little Rock, Arkansas 72204-1099

(Received 14 January 1998; accepted for publication 4 September 1998)

Sentences were processed through simulations of cochlear-implant signal processors with 6, 8, 12, 16, and 20 channels and were presented to normal-hearing listeners at +2 dB S/N and at -2 dB S/N. The signal-processing operations included bandpass filtering, rectification, and smoothing of the signal in each band, estimation of the rms energy of the signal in each band (computed every 4 ms), and generation of sinusoids with frequencies equal to the center frequencies of the bands and amplitudes equal to the rms levels in each band. The sinusoids were summed and presented to listeners for identification. At issue was the number of channels necessary to reach maximum performance on tests of sentence understanding. At +2 dB S/N, the performance maximum was reached with 12 channels of stimulation. At -2 dB S/N, the performance maximum was reached with 20 channels of stimulation. These results, in combination with the outcome that in quiet, asymptotic performance is reached with five channels of stimulation, demonstrate that more channels are needed in noise than in quiet to reach a high level of sentence understanding and that, as the S/N becomes poorer, more channels are needed to achieve a given level of performance.

© 1998 Acoustical Society of America. [S0001-4966(98)04912-1]

PACS numbers: 43.71.Es, 43.71.Ky, 43.66.Ts [WS]

INTRODUCTION

Cochlear implants currently use a minimum of 4 and a maximum of 24 electrodes. It is reasonable to assume that the greater the number of electrodes, the better the speech understanding. This, however, does not seem to be the case. Patients with 6-channel processors achieve scores on tests of speech understanding which are similar to the scores of patients using processors with many more channels (see Dorman *et al.*, in press; Fishman, Shannon, and Slattery, 1997; Wilson, 1997; Tyler *et al.*, 1996). Thus, the data from patients leave unanswered the question of how many electrodes, or channels, should be implemented in a cochlear implant.

Another approach to answering this question is to test normal-hearing listeners with signals which have been processed in the manner of a cochlear-implant signal processor. By using normal-hearing listeners as subjects, many sources of variability inherent in the testing of cochlear-implant patients, such as differences in electrical stimulation strategy, differences in the survival of cell bodies in the spiral ganglion, and differences in the location of electrodes relative to remaining neural tissue, are eliminated. Of course, "simulations" have their own set of limitations, the most notable being that auditory stimulation cannot replicate the mostly

unknown dynamics of current spread in the cochlea. Nonetheless, experiments with normal-hearing listeners give insight into the number of channels which, in the best case, would be needed to reproduce speech adequately for cochlear-implant patients.

In simulation experiments with normal-hearing listeners, signals are typically bandpass filtered into n channels and are rectified and smoothed. The root mean square (rms) energy of the envelope in each channel is computed and signals, either sinusoids with frequencies equal to the center frequencies of the bandpass filters, or noise bands the width of the analysis filters, are output to listeners. When processed in this manner, sentences, *in quiet*, can be understood with greater than 90% accuracy when as few as four channels are implemented (Shannon *et al.*, 1995; Dorman *et al.*, 1997). These data suggest that if electrical stimulation could reproduce the stimulation produced on the basilar membrane by a small number of fixed-frequency sine waves, or noise bands, then only a small number of channels would need to be implemented in a cochlear-implant signal processor to achieve a very high level of sentence understanding.

The results of experiments on sentence understanding in quiet most likely underestimate the number of channels necessary to achieve a high level of sentence understanding in

noisy, “real world” situations. If this were the case, then the view that only a small number of channels are necessary in a cochlear-implant signal processor would need to be modified. To determine whether this is the case, in the present experiment, sentences were processed into 6, 8, 12, 16, and 20 channels and were presented to normal-hearing listeners against a background of speech-shaped noise at +2 dB S/N and -2 dB S/N. The +2 dB S/N ratio was chosen because pilot experiments suggested that interpretation of the results would not be confounded by a ceiling effect on performance. The -2 dB S/N was chosen to assess whether the number of channels necessary to reach maximum performance increased as the signal to noise ratio decreased.

I. METHOD

A. Subjects

Twenty-one normal-hearing students at Arizona State University, who ranged in age from 22 to 55 years, participated in the experiment. The subjects were divided into two groups (of 11 and 10 subjects) for testing.

B. Speech material

The test signals were sentences taken from the H.I.N.T. lists (Nilsson, Soli, and Sullivan, 1994). Each sentence was composed of between four and seven words. Each word was counted in the calculation of percent-correct scores. The sentences were presented at +2 and -2 dB S/N. Twenty sentences were used in each of the five test conditions (i.e., for processors with 6, 8, 12, 16, and 20 channels) for each of the S/N ratios. The noise was spectrally shaped based on the averaged spectrum of the H.I.N.T. sentences (Nilsson *et al.*, 1994). For each S/N condition, the stimuli from all five channel conditions were randomized into a single test sequence. All test materials were stored on computer disk and were output via custom software routines using MATLAB (The MathWorks, Inc.) software and a 16-bit D/A converter.

C. Signal processing

A software version of a cochlear-implant signal processor, similar to that of the Med El Combi-40, was implemented using the MATLAB signal-processing toolbox. Signal processing was implemented in the following manner. Signals were first processed through a pre-emphasis filter (low pass below 1200 Hz, -6 dB per octave) and then bandpassed into n frequency bands (where n varied from 6 to 20) using sixth-order Butterworth filters. The envelope of the signal was extracted, using a 4-ms rectangular window, by full-wave rectification and low-pass filtering (second-order Butterworth) with a 400-Hz cutoff frequency. [A 400-Hz cutoff frequency was used to conform to the cutoff frequency commonly used in the Combi-40’s signal processor. Shannon *et al.* (1995) found no difference in patient performance with low-pass filters set at 160 Hz and above.] Sinusoids were generated with amplitudes equal to the root-mean-square (rms) energy of the envelopes (computed every 4 ms) and frequencies equal to the center frequencies of the bandpass

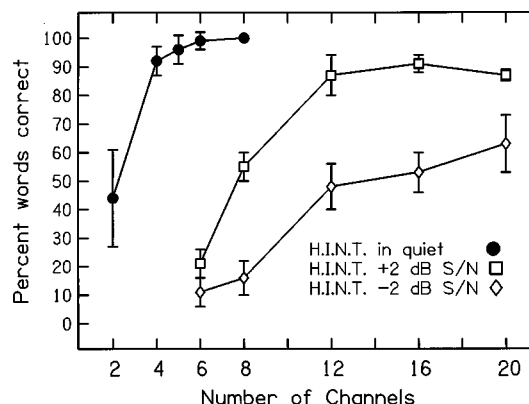


FIG. 1. Sentence intelligibility as a function of the number of channels of stimulation. The parameter is the signal-to-noise ratio. Error bars indicate ± 1 standard deviation.

filters. The sinusoids of each band were summed and presented to the listeners at 72 dB SPL (re: vowel peaks) through Sennheiser HMD 410 headphones.

Signal processors with 6, 8, 12, 16, and 20 channels were constructed. The center frequencies of the channels in the processors with 6 and 8 channels were equally spaced on a logarithmic scale. The center frequencies for the 6-channel processor were 394, 639, 1038, 1685, 2736, and 4444 Hz. For the 8-channel processor, the center frequencies were 366, 526, 757, 1089, 1566, 2253, 3241, and 4662 Hz. For processors with 12, 16, and 20 channels, the channel center frequencies were computed according to the equation: $1100 \log(f/800 + 1)$, where f is the frequency in Hz [this is similar to the technical mel scale of Fant (1973)]. The center frequencies for the 12-channel processor were 274, 453, 662, 905, 1190, 1521, 1908, 2359, 2885, 3499, 4215, and 5050 Hz. The center frequencies for the 16-channel processor were 216, 343, 486, 647, 828, 1031, 1260, 1518, 1808, 2134, 2501, 2914, 3378, 3901, 4489, and 5150 Hz. The center frequencies for the 20-channel processor were 182, 280, 388, 507, 638, 782, 940, 1114, 1306, 1517, 1749, 2004, 2284, 2593, 2932, 3306, 3717, 4169, 4666, and 5213 Hz.

D. Procedures

The subjects were tested in two groups. One group was tested with the sentence material at +2 dB S/N, and the other at -2 dB S/N. For practice, at each S/N the listeners were presented with 100 sentences from the TIMIT sentence database (Lamel *et al.*, 1986). Five blocks of 20 sentences were presented during practice. In each block the sentences were processed in the manner of one of the channel conditions (i.e., $n=6, 8, 12, 16, \text{ and } 20$). The words in each sentence were displayed on a computer screen following the presentation of the sentence. During the test, sentences were presented once and responses were entered, either by the subject, or by the third author, by typing on the computer keyboard. The test was self paced.

II. RESULTS

The results for sentences presented at +2 dB S/N are shown as the middle function in Fig. 1. The mean scores for

processors with 6, 8, 12, 16, and 20 channels were 21, 55, 87, 91, and 87 percent correct, respectively. A repeated-measures analysis of variance demonstrated a significant main effect for channels ($F_{4,10}=418$, $p<0.000\ 001$). *Post hoc* tests according to Scheffe, using an alpha of 0.05, indicated that eight channels allowed significantly better performance than six channels, and that 12 channels allowed significantly better performance than eight channels. Performance reached asymptote with 12 channels of stimulation, i.e., performance with 12-, 16-, and 20-channel processors did not differ.

The results for sentences presented at -2 dB S/N are shown as the right-most function in Fig. 1. The mean scores for processors with 6, 8, 12, 16, and 20 channels were 11, 16, 48, 53, and 62 percent correct, respectively. A repeated-measures analysis of variance demonstrated a significant main effect for channels ($F_{4,9}=100.0$, $p<0.000\ 01$). *Post hoc* tests according to Scheffe, using an alpha of 0.05, indicated that performance in the 6- and 8-channel conditions differed from that in the 12-, 16-, and 20-channel conditions, that performance in the 12-, and 16-channel conditions differed from that in the 6-, 8-, and 20-channel conditions, and that performance in the 20-channel condition differed from performance in all of the other conditions.

III. DISCUSSION

In the introduction we noted that, in quiet, only four channels of stimulation are necessary to allow sentences to be understood with greater than 90% accuracy. This is not the case in noise. At $+2$ dB S/N, where performance is not constrained by a ceiling effect, 12 channels are necessary to reach maximum performance. At -2 dB S/N, 20 channels are necessary to reach maximum performance. As the S/N ratio becomes poorer, more channels are needed to achieve a given level of performance. These findings are illustrated in Fig. 1, where the current data and data on the recognition of the H.I.N.T. sentences in quiet, from Dorman *et al.* (1997), are plotted. The 8-channel processor, which allowed a mean score of 100-percent correct in quiet, allowed a mean score of 55-percent correct at $+2$ dB S/N and allowed a mean score of 16-percent correct at -2 dB S/N. If speech in noise is to be understood, then signal processors for cochlear implants should have more than a few channels and, most generally, the more channels, the better.

Achieving the goal of transmitting information to implant patients through a large number of channels has proven difficult. For example, when the number of electrodes activated in a 22-electrode array has been systematically reduced and speech understanding has been measured in quiet, six or seven electrodes have provided the same intelligibility as 20 electrodes (Fishman *et al.*, 1997; Wilson, 1997). Or, put another way, increasing the number of electrodes beyond six or seven has not improved intelligibility. This outcome suggests that it may be some time before 20 independent, information-bearing channels can be realized for cochlear implant patients. If this is the case, then the goal of 12 chan-

nels is a reasonable one. As shown in Fig. 1, sentence intelligibility functions at $+2$ and -2 dB S/N show large increases in performance when the number of channels increases from 8 to 12, but show no significant improvement from 12 to 16. The large improvement in performance for the 12-channel processor relative to the 8-channel processor is due, most likely, to the increase in the number of filters, from three to five, allocated to the frequency domain of F_2 (i.e., between 900 and 2500 Hz). The increase in number of filters allows better resolution of F_2 frequency. Other factors which may underlie the improvement in performance for the 12-channel processor relative to the 8-channel processor are better spectral contrast within each channel and an increase in the number of channels with a high S/N. Both of these factors are a consequence of the decrease in filter width in the 12-channel condition.

If providing 12 channels of stimulation proves beyond the reach of current technology, then even a small increase in the number of effective channels of stimulation from current devices would be of great benefit. As shown in Fig. 1, at $+2$ dB S/N a 34-point improvement in intelligibility was obtained when the number of channels was increased from six to eight. In an unpublished experiment, we have found a 20-point improvement in intelligibility at $+6$ dB S/N when the number of channels was increased from six to eight. Thus, current implant devices need to be improved only a little, in terms of the effective channels of stimulation, in order to provide a large improvement in patient performance in noise.

ACKNOWLEDGMENT

This research was supported by grant No. DC RO 1 00654-7 from the National Institute on Deafness and Other Communication Disorders.

- Dorman, M. F., Loizou, P., and 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," *J. Acoust. Soc. Am.* **102**, 2403–2411.
- Dorman, M., Loizou, P., Fitzke, J., and Tu, Z. (in press). "The recognition of NU-6 words by cochlear-implant patients and by normal-hearing subjects listening to NU-6 words processed in the manner of CIS and SPEAK strategies," *Ann. Otol. Rhinol. Laryngol.*
- Fant, G. (1973). *Speech Sounds and Features* (MIT, Cambridge, MA).
- Fishman, K., Shannon, B., and Slattery, W. (1997). "Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor," *J. Speech Hear. Res.* **40**(5), 1201–1215.
- Lamel, L., Kassel, R., and Seneff, S. (1986). "Speech database development: Design and analysis of the acoustic-phonetic corpus," *Proceedings of DARPA Speech Recognition Workshop*, pp. 100–109.
- Nilsson, M., Soli, S., and Sullivan, J. (1994). "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and noise," *J. Acoust. Soc. Am.* **95**, 1085–1099.
- Shannon, R., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Tyler, R., Gantz, B., Woodworth, G., Parkinson, A., Lowder, M., and Schum, L. (1996). "Initial independent results with the Clarion cochlear implant," *Ear and Hearing* **17**(6), 528–536.
- Wilson, B. (1997). "The future of cochlear implants," *British Journal of Audiology* **31**, 205–225.