Title: The recognition of sentences in noise by normal-hearing listeners using simulations of SPEAK-type cochlear implant signal processors.

Running head: Intelligibility of SPEAK-type processors

Authors: Philipos C. Loizou, Ph.D., Department of Applied Science, University of Arkansas at Little Rock, Little Rock, AR 72204; Michael F. Dorman, Ph.D., Department of Speech and Hearing Science, Arizona State University, Tempe, AZ 85287-0102 and Division of Otolaryngology, University of Utah Health Sciences Center; Zhemin Tu, M.S., Department of Applied Science, University of Arkansas at Little Rock, Little Rock, AR 72204; Jeanette Fitzke, M.A., Department of Speech and Hearing Science, Arizona State University, Tempe, AZ 85287-0102.

Grant support: This work was supported by NIH grants R01 DC-03421 and R01 DC-000654 from NIDCD.

Reprint requests and correspondence to:
Philipos C. Loizou, Ph.D., Department of Applied Science, University of Arkansas at Little Rock, Little Rock, AR 72204. Telephone 501-569-8067; Fax 501-569-8020
E-mail: loizou@ualr.edu

Abstract

To assess whether more channels are needed to understand speech in noise than in quiet, speech was processed in a manner similar to SPEAK-like cochlear implant processors and presented at +2 dB S/N to normal hearing listeners for identification. The number of analysis filters varied from 8 to 16, and the number of maximum channel amplitudes selected in each cycle varied from 2 to 16. The results show that more channels are needed to understand speech in noise than in quiet, and that high levels of speech understanding can be achieved with 12 channels. Selecting more than 12 channel amplitudes out of 16 channels did not yield significant improvements in recognition performance.

Introduction

Previous studies[1,2] have shown that, when speech is processed in the manner of cochlear implants and presented in quiet to normal hearing listeners, greater than 90% sentence recognition can be achieved with as few as four channels of stimulation. In the real world, however, implant patients often need to communicate in noisy environments. At issue is whether more channels are needed in noisy environments to achieve high levels of speech understanding. To gain insight at this issue, speech was processed through a simulation of a SPEAK-like cochlear implant processor and presented at +2 dB S/N to normal-hearing listeners. The total number of channels was varied from 8 to 16. To investigate whether the SPEAK-type implant processors, which select in each cycle a small number of channel outputs from a larger set, offer any advantages in noise, we also varied the number of channels selected from 2 to 16.

Method

Subjects. Ten normal-hearing subjects at Arizona State University, ranging in age from 22 to 55 years, participated in the experiment. The subjects were paid for their participation.

Speech material. Sentences taken from the H.I.N.T database[3] were used. The sentences were presented at +2 dB S/N, using spectrally-shaped noise computed from the long-term average spectrum of speech. Twenty sentences were used for each test condition. All sentences were stored on computer disk and were output via custom software routines using MATLAB software and a 16 bit D/A converter.

Signal Processing. In SPEAK-type or “n-of-m” processors, the signal is bandpassed into *m* channels, and the *n* ($n<m$) channels with the largest energy are selected for stimulation in each cycle. The sentences were processed in a manner similar to a SPEAK cochlear implant processor[4] to produce three types of processors: n-of-8 (n=2,4,6,8), n-of-12 (n=2,4,6,8,12), and n-of-16 (n=2,4,6,8,12,16) processors. The signal was first processed through a pre-emphasis filter (low-pass below 1200 Hz, 6 dB/octave) and then bandpassed into 8, 12 or 16 frequency bands using sixth-order Butterworth filters. Logarithmic spacing was used for the n-of-8 processors, and mel spacing was used for the n-of-12 and n-of-16 processors. The envelope of the signal was then 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 *n* maximum envelopes (computed every 4 msecs) and frequencies equal to the center frequencies of the *n* selected channels. The *n* sinusoids were finally summed and presented to normal-hearing listeners for identification through Sennheiser HMD 410 headphones. The number of channels (*n)* selected varied from 2 to 16 depending on the processor. For the 4-of-12 processor, for example, *n*=4, and in each cycle the four (out of 12) maximum channel outputs with the largest amplitude were selected for stimulation.

Procedure. The subjects were tested with the n-of-8, n-of-12 and n-of-16 processors using the HINT sentences in +2 dB S/N. Twenty sentences were used for each test condition. The subjects were presented for practice with 100 TIMIT sentences processed in a manner similar to each processor condition. The test sentences within each processor condition were all randomized. During the test, the responses were entered by typing on the computer keyboard. The test was self-paced.

Results

The results, in terms of percent words correct, are shown in Figure 1. The sentence recognition performance for the n-of-8 processors is shown in Figure 1a. The mean word scores for 2-, 4-, 6-, and 8-of-8 processors were 38%, 50%, 60% and 55% correct respectively. Repeated measures analysis of variance indicated a significant main effect [$F(3,27)=29.807$, $p<0.0001$] for the number of channels selected out of 8 channels. Post-hoc tests showed that the mean word score obtained with the 8-of-8 processor was significantly lower [$t(9)=2.414$, $p<0.05$] than the mean word score obtained with the 6-of-8 processor. Seven out of the 10 listeners showed a decrease in performance from the 8-of-8 processor to the 6-of-8 processor. The sentence recognition performance for the n-of-12 processors is shown in Figure 1b. The mean word scores for 2-, 4-, 6-, 8- and 12-of-12 processors were 46%, 61%, 83%, 90% and 87% correct respectively. Repeated measures analysis of variance indicated a significant main effect [$F(4,36)=94.216$, $p<0.0001$] for the number of channels selected out of 12 channels. Finally, the results for the n-of-16 processors is shown in Figure 1c. The mean word scores for 2-, 4-, 6-, 8-, 12- and 16-of-16 processors were 36%, 69%, 77%, 84%, 94% and 92% correct respectively.

Repeated measures analysis of variance indicated a significant main effect [F(5,45)=82.997, p<0.0001] for the number of channels selected out of 16 channels. Post-hoc tests showed no significant difference between the scores obtained with 12-of-16 processor and 16-of-16 processor.

Discussion

Our previous studies[1] showed that in quiet only four or five channels of stimulation were necessary to achieve greater than 90% accuracy in sentence understanding. The present study shows that this is not the case in noise. At +2 dB S/N, where performance is not constrained by a ceiling effect, speech understanding improves significantly as more channels are added. Twelve channels were needed to reach maximum performance. Using 16 channels did not produce any significant improvements in sentence understanding compared to the 8-of-12 processor (92% vs. 90% correct). The results of this study also show that if a small number (<12) of channels is used, then a SPEAK-like or "n-of-m" processor might offer some advantages in noise compared to a fixed-channel processor (i.e., an m-of-m processor). The mean word score obtained with a processor that selected, in each cycle, six out of eight channels (6-of-8 processor) was significantly higher than the mean word score obtained with a processor that selected all channels (8-of-8 processor) for stimulation. Seven out of the 10 listeners showed a decrease in performance listening to speech processed through an 8-of-8 processor versus a 6-of-8 processor. These results are consistent with those reported by Kendall *et al.*[5]. Overall, our results suggest that cochlear implant patients might gain more benefit in noise using an n-of-m strategy (e.g., such as the one currently implemented in the Med-El/COMBI 40+ device) rather than a fixed-channel strategy (i.e., an m-of-m strategy).

Acknowledgements

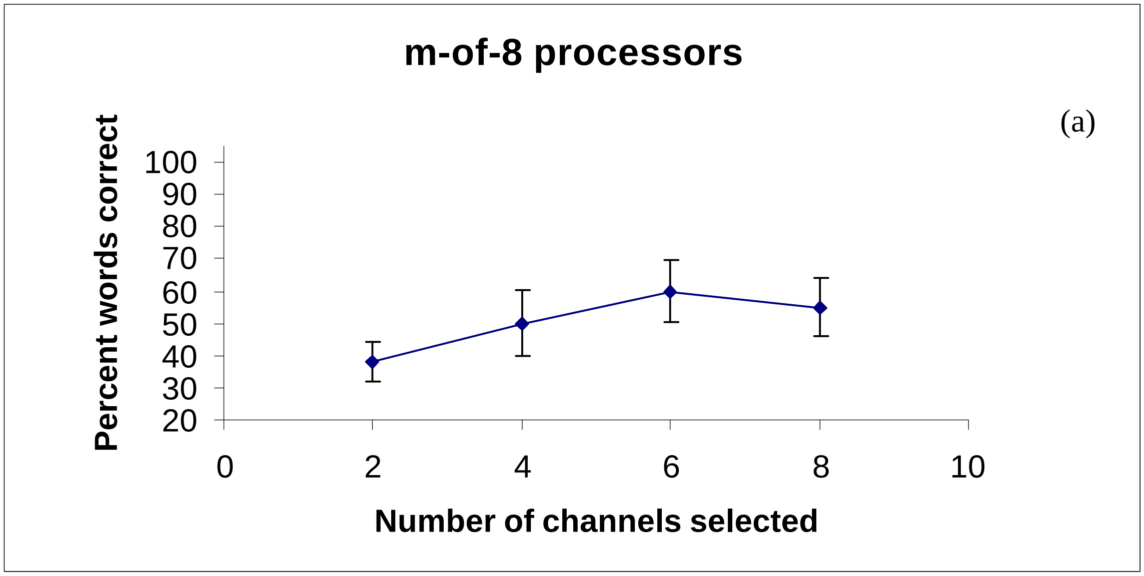
This work was supported by NIH grants R01 DC-03421 and R01 DC-000654 from NIDCD.

References

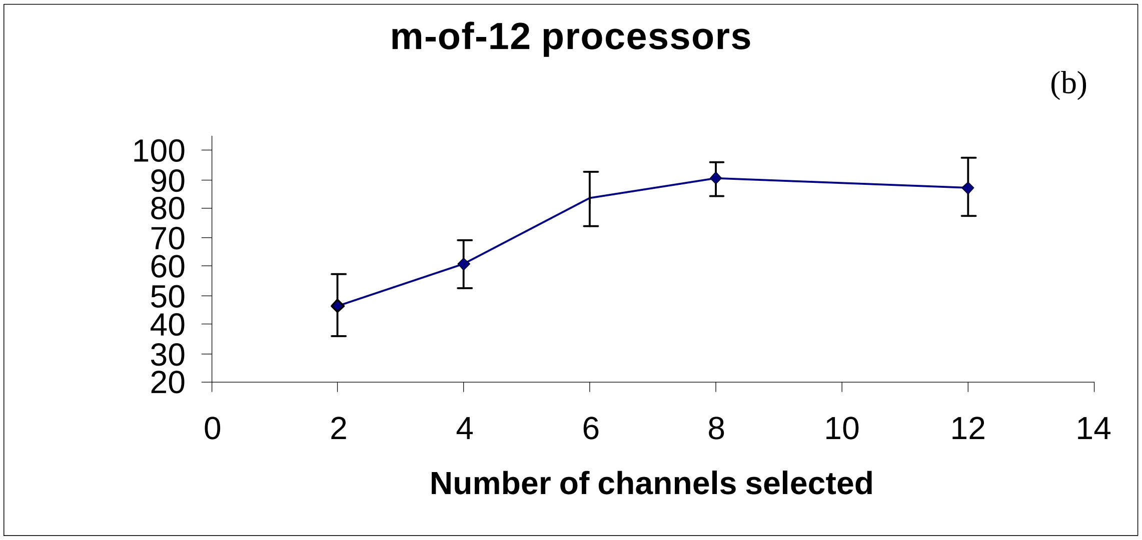
1. Dorman M., Loizou P., and Rainey D. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs, J. Acoust. Soc. Am. 1997, 102(4):2403-2411.
2. Shannon R., Zeng F-G, Kamath V., Wygonski J. and Ekelid M. Speech recognition with primarily temporal cues, Science 1995, 270:303-304.
3. Nilsson M., Soli, S. and Sullivan, J., Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise, J. Acoust. Soc. Am. 1994, 95(2):1085-1099.
4. Loizou, P. Mimicking the human ear: An overview of signal-processing techniques for converting sound into electrical signals in cochlear implants, IEEE Signal Processing Magazine 1998, 15(5):101-130.
5. Kendall M., Summerfield Q. and Chambers J. Simulation of a multi-channel cochlear implant: Perceptual consequences of strategies for channel selection, 1997 Conference on Implantable Auditory Prostheses 1997, 83.

Figure Captions

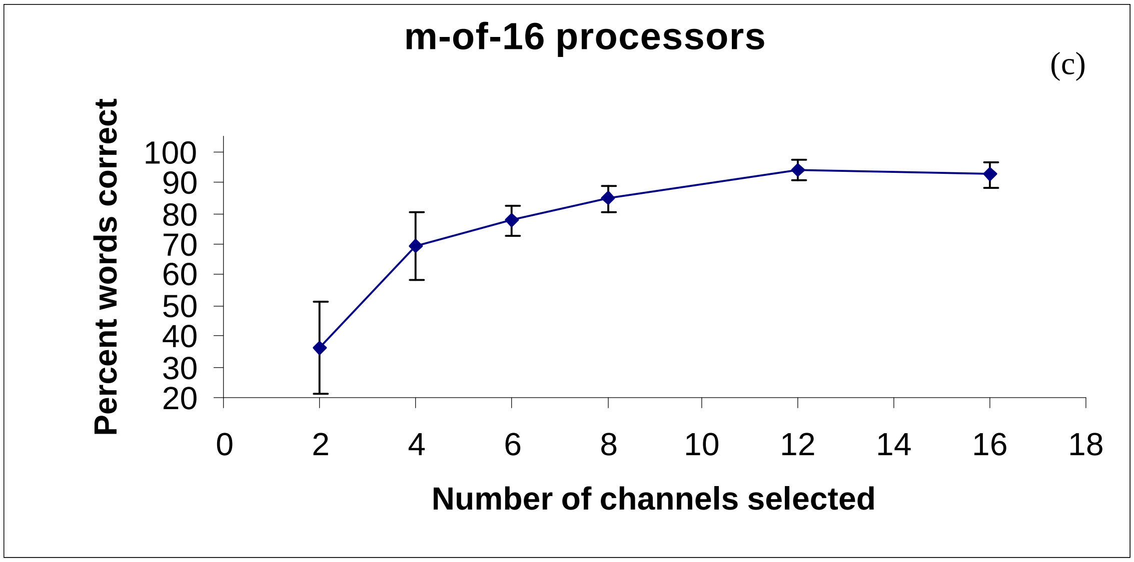
Figure 1. Speech recognition as a function of the number of channels selected. Error bars indicate ±1 standard deviations.


m-of-8 processors

(a)


m-of-12 processors

(b)


m-of-16 processors

(c)