

Performance of spectral-maxima and fixed-channel algorithms for coding speech in quiet and in noise

Michael F. Dorman, Philip C. Loizou, Tony Spahr and Erin Maloff

Abstract—Sentence intelligibility in quiet and in noise was assessed for two types of signal processing algorithms commonly implemented for cochlear implants. Experiment 1 determined the number of m channels in an m -of-18 processor that are necessary for asymptotic performance. Experiment 2 determined the number of fixed channels necessary to equal the performance of the best spectral-maxima processor. A 4-of-18 processor produced asymptotic performance in quiet. An 8-of-18 processor produced asymptotic performance in noise. A fixed channel processor with 10 channels allowed the same performance as the best performing, spectral-maxima processor. Given equivalent levels of performance, fixed channel processors offer an advantage in simplicity over spectral-maxima processors for cochlear implants.

Index terms—cochlear implants, speech processing

I. INTRODUCTION

Current cochlear implants employ two, quite different, processing strategies for speech analysis and transmission [1]. One strategy divides the input signal into 20 channels, estimates the energy in each channel and outputs a pulse, with amplitude proportional to the energy in each channel, from the 6-10 channels with the highest (maximum) energy. This spectral-maxima scheme has historical antecedents in the peak-picking channel vocoders of the 1950s [2] and the Haskins Laboratories' Pattern Playback [3]. A spectral maxima strategy is efficient for speech transmission because, by picking the channels with the highest energy, the peaks in the short-term spectrum will be transmitted. This strategy is also referred to as the m -of- n strategy, where m indicates the number of spectral maxima selected out of a total of n channels. Such a strategy has been employed in sinusoidal coders (using the FFT) for low bit rate speech

coding [4], and is also being used in the Nucleus cochlear implant device [1][5].

The second strategy, termed a *fixed-channel* strategy, divides the input signal into a small number of channels, e.g. 6-12, estimates the energy in each channel, and outputs a pulse, with amplitude proportional to the energy in each channel, to each of the channels. It is called the fixed-channel strategy, because the same channels are used in each cycle, compared to the "m-of-n" strategy in which only m out of n channels are used in each cycle, and the m channels selected can vary from cycle to cycle. This strategy has historical antecedents in the fixed-channel vocoders of the 1950s [2], and is currently being used in the Clarion, Med-El and Nucleus-24 implant devices. Both strategies allow very high levels of sentence understanding by deaf adults and congenitally deaf children [1].

At issue, in the present study, is whether one of the strategies has an inherent advantage in coding speech in quiet and/or in noise. The answer has important implications for signal processor design for cochlear implants. To answer this question we processed speech in the manner of spectral-maxima and fixed-channel processors and output the signals as the sum of a set of sine waves to normal-hearing individuals. Our aims were, (i) to determine the number of 'm' channels that need to be picked in an 'm-of-n' processors in quiet and in noise to reach asymptotic performance, and (ii) to compare the performance of spectral-maxima processors and fixed channel processors in quiet and in noise.

II. EXPERIMENT 1: PERFORMANCE OF THE SPECTRAL-MAXIMA SPEECH STRATEGY

A. Signal processing.

Nine spectral-maxima processors, or "m-of-n" processors, were implemented. For these processors n was fixed at 18 channels and m varied from 2 to 16 in two channel steps. For all of the processors, signals were first processed through a pre-emphasis filter (low-pass below 1200 Hz, 6dB/octave) and then bandpassed into 18 frequency bands ($n=18$) using sixth-order Butterworth filters and mel spacing.

M. Dorman, T. Spahr and E. Maloff are with the Department of Speech and Hearing Sciences, Arizona State University, and P. Loizou is with the Dept. of Electrical Engineering, University of Texas-Dallas, Richardson, TX 75083-0688 (loizou@utdallas.edu).

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 m channels with the highest energy in each 4-msec update cycle and with frequencies equal to the center frequencies of the m selected channels. The m sinusoids were finally summed and presented to normal-hearing listeners for identification through Sennheiser HMD 410 headphones.

B. Speech material.

One hundred twenty sentences from the TIMIT database were used to assess performance. Fifteen sentences were used in each of the 8 test conditions. The signals were presented in quiet and in speech-shaped noise at +6 dB S/N. The signal to-noise-ratio was chosen, following pilot experiments, to both avoid a ceiling effect and to allow a reasonably high level of performance. The same sentences were used for the quiet and noise conditions.

C. Subjects.

Twenty subjects were tested, 10 in the quiet condition, and 10 in the noise condition.

D. Results

Percent correct words as function of the number of channels for signals presented in quiet and in noise are shown in Figure 1. Repeated measures ANOVAs revealed a main effect for number of channels in both the quiet and noise conditions (quiet: $F_{7,63}=207, p<.00001$; noise; $F_{7,63}=192, p<.00001$). Post-hoc tests indicated that performance reached asymptote with 4 channels of stimulation in quiet and with 8 channels of stimulation in noise.

E. Discussion

The aim in Experiment 1 was to determine the number of m channels that need to be picked in an m-of-20 processing scheme to achieve asymptotic performance in quiet and in noise. In quiet, where performance is constrained by a ceiling effect, picking 4 of 20 channels was sufficient to reach greater than 90 % correct. In quiet, then, signal processors for cochlear implants could be implemented with relatively few output channels. However, in a modest amount of noise, i.e., at +6 dB S/N, 8 channels are necessary for asymptotic performance. Since a +6 dB S/N is not a particularly poor signal to noise ratio, it is reasonable to suggest that spectral-maxima algorithms for cochlear implants use at least 8 or more m channels for every day patient use.

III. EXPERIMENT 2: PERFORMANCE OF THE FIXED-CHANNEL SPEECH STRATEGY

The aim of experiment 2 was to determine the number of channels necessary in fixed-channel processors to equal the performance of spectral-maxima processors in quiet and in noise.

A. Signal processing.

Fixed-channel processors were implemented with 4, 6, 8, 10, 12, 16 and 20 channels. Signal processing was identical to that for the spectral-maxima processors except that on every update cycle sinusoids were output to all of the channels in the processor. Three spectral-maxima processors were also implemented. For these processors $m=3,6,9$ and $n=20$.

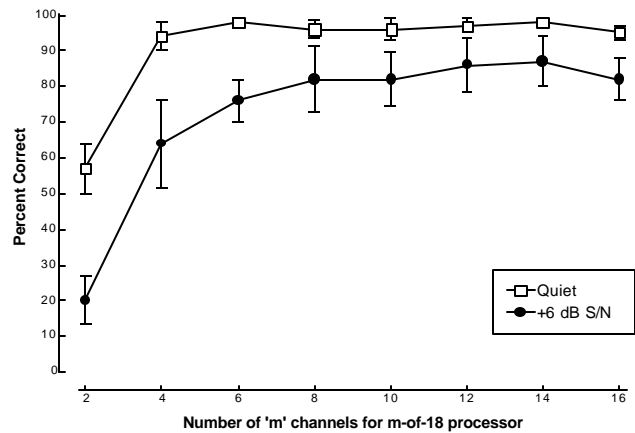


Figure 1. The intelligibility of words in the TIMIT sentences in quiet and in noise as a function of the number of m channels selected out of 18 channels.

The fixed-channel technique is similar to Dudley's channel vocoder [6]. The main difference is that in Dudley's vocoder system, speech was synthesized by exciting a bank of bandpass filters with quasi periodic pulses when speech was voiced, and with white noise when speech was unvoiced. In the fixed-channel technique, speech (voiced and unvoiced) is synthesized as a sum of a small number of sinewaves. No voiced/unvoiced detector or pitch detector is needed. More information about this technique can be found in [7].

B. Test material.

Ten sentences from the HINT database [8] were used in each processor condition. The signals were presented in quiet and at 0 dB S/N. This signal to noise ratio (0 dB) was chosen, following pilot experiments, to provide a level of performance near that of the noise condition in Experiment 1. Because the sentence material was less difficult than in Experiment 1, the signal to noise ratio was reduced to 0 dB.

C. Results

Intelligibility as a function of the number channels is shown in the left panel of Figure 2. Intelligibility as a function of the number of m channels in an m -of-20 processor is shown in the right panel. Repeated measures ANOVAs for the fixed channel condition indicated a main effect for number of channels in quiet ($F_{6,63}=10.6$, $p.<.0001$) and in noise ($F_{6,63}=179.8$, $p.<.000001$). Post-hoc tests indicated a performance asymptote for the quiet condition at 4 channels and an asymptote for the noise condition at 10 channels.

IV. DISCUSSION

The aim of Experiment 2 was to determine the number of channels in a fixed-channel processor that are needed to produce scores equivalent to those of a spectral-maxima processor. In quiet, a processor with 6 fixed channels allowed the same level of performance (99 % correct) as allowed by a 9-of-20 processor (99 % correct). Of course, scores in both cases were constrained by a ceiling effect. In the case when performance was not constrained by a ceiling effect, i.e., the 0 dB S/N condition, a processor with 10 fixed channels allowed the same level of performance (73 % correct) as a 9-of-20 processor (75 % correct). Overall, the data suggest that there is no inherent advantage, in either quiet or in moderate amounts of noise, of using a spectral-maxima processor in contrast to a fixed-channel processor.

This outcome has two implications for cochlear implants. One is that implant processors can use fixed-channel algorithms and avoid the complexity of spectral-maxima algorithms. The second is that if implant patients who are tested with both fixed channel and spectral-maxima algorithms perform better with the spectral-maxima algorithm, then the difference in performance is probably not due to the inherent superiority of the spectral-maxima algorithm in transmitting speech information. Rather, the superiority must be due to other factors, including a possible reduction in current interaction with the spectral-maxima scheme.

Acknowledgements

This research was supported by grants from the National Institute on Deafness and other Communication Disorders/NIH to MFD (RO1 DC-000654-9) and to PCL (RO1 DC- 03421).

References

- [1] P. Loizou, "Mimicking the human ear," *IEEE Signal Processing Magazine*, vol. 15, no. 5, pp.101-130, September 1998.
- [2] M. Schroeder, "Vocoders: analysis and synthesis of speech," *Proc. of IEEE*, vol. 54, pp. 720-734, May 1966.
- [3] F. Cooper, A. Liberman and J. Borst, "The interconversion of audible and visible patterns as a basis for research in the

perception of speech," *Proceedings of the National Academy of Sciences*, vol. 37, pp. 318-322, 1951.

- [4] P. Loizou and A. Spanias, "Vector quantization of principal spectral components," *Proc. of 24th Asilomar Conference on Signals, Systems and Computers*, pp. 654-658, 1990.
- [5] H. McDermott, C. McKay, and A. Vandali, "A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant," *Journal of the Acoustical Society of America*, vol. 91, pp. 3367-3371, 1992.
- [6] Dudley, H., "Remaking speech," *Journal of Acoustical Society of America*, vol. 11, pp. 169-177, 1939.
- [7] P. Loizou, M. Dorman, Z. Tu, "On the number of channels needed to understand speech," *Journal of Acoustical Society of America*, vol. 106, no. 4, pp. 2097-2103, 1999.
- [8] M. Nilsson, S. Soli and J. Sullivan, "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," *Journal of Acoustical Society of America*, vol. 95, no. 2, pp. 1085-1099, 1994.

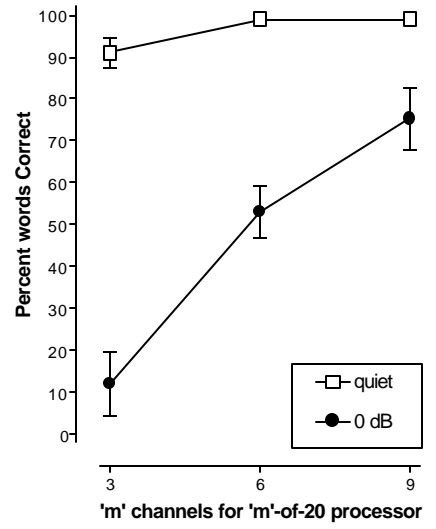
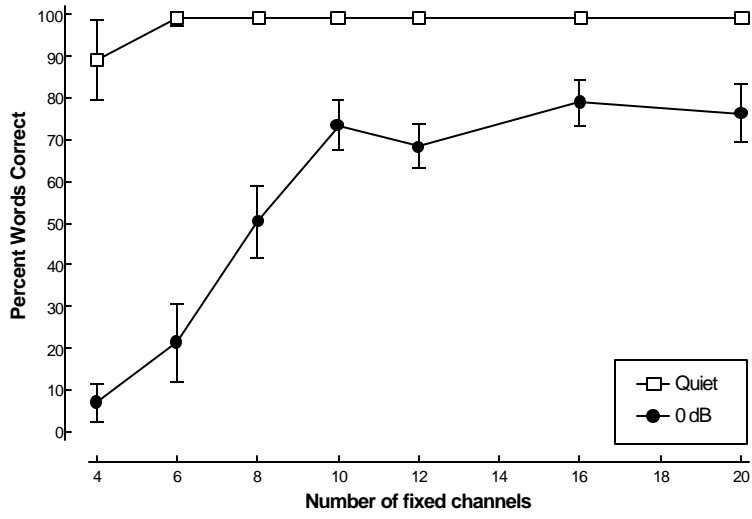


Figure 2. In the left panel, the intelligibility of words in the HINT sentences in quiet and in noise as a function of the number of fixed channels is shown. In the right panel, word intelligibility is shown when $m=3,6,9$ for an m -of-20 processor.