

Improving Consonant Intelligibility for Ineraid Patients Fit with Continuous Interleaved Sampling (CIS) Processors by Enhancing Contrast Among Channel Outputs

Dorman, Michael F.; Loizou, Philipos C.

Arizona State University, Tempe, Arizona, and University of Utah Health Sciences Center (M.F.D.), Salt Lake City, Utah; and University of Arkansas at Little Rock (P.C.L.), Little Rock, Arkansas

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

Received December 5, 1995; accepted

Abstract [TOP](#)

Objective: In Experiment 1 the objective was to determine whether patients who have been implanted with the Ineraid electrode array perform better on tests of consonant identification when signals are processed through a continuous interleaved sampling (CIS) processor than when signals are processed through an analogue (Ineraid) processor. In Experiment 2 the objective was to determine, for patients using the CIS strategy, whether identification accuracy for stop consonant place of articulation could be improved by enhancing differences in the patterns of the signal processor channel outputs.

Design: In Experiment 1, 16 consonants were presented in VCV format for identification. In Experiment 1 the CIS patients evidenced difficulty in identifying /p t k/. Therefore, in Experiment 2 the voiceless stop consonants were presented in two stimulus conditions. In one, the stimuli were unfiltered. In the other, the stimuli were individually filtered so as to enhance the differences in channel outputs for /p/, /t/, and /k/.

Results: In Experiment 1 the patients performed better with CIS processors than with analogue processors. In Experiment 2 the "enhanced" stimuli were identified with better accuracy than were the unfiltered stimuli.

Conclusions: We confirm that Ineraid patients achieve higher scores on tests of consonant identification when using a CIS processor than when using an analogue processor. Errors in identification of stop consonant place of articulation, when using a CIS processor, are due to the similarity in the patterns of the processor's channel outputs. By showing that consonant intelligibility can be improved by filtering, we show that we have not reached the limit of speech understanding that can be supported by the population of neural elements remaining in our patients' auditory systems.

In [1991 Wilson, Finley, Lawson, Wolford, Eddington, and Rabinowitz](#) described a new signal processing strategy, continuous interleaved sampling (CIS), for cochlear implants (see, also, [Lawson, Wilson, & Finley, 1993](#)). The CIS strategy was tested, first, on patients who used the Ineraid cochlear implant because CIS processors could be coupled directly to intracochlear electrodes via the Ineraid's percutaneous pedestal. In contrast to the Ineraid, which presents analogue waveforms simultaneously to four electrodes, a CIS processor provides continuous, high-rate, pulsatile stimulation in a nonoverlapping sequence to six electrodes (for Ineraid patients). Because the CIS processor can provide more channels of stimulation than the Ineraid, and because channel interactions should be reduced by virtue of the nonsimultaneous pulsatile stimulation, frequency discrimination and, therefore, speech understanding should be better with the CIS processor than with the Ineraid processor. Results that are consistent with this speculation have been reported by [Dorman \(1994\)](#) and [Dorman, Smith, Smith,](#)

Article Outline

[Abstract](#)

[Experiment 1](#)

○ [Method](#)

○ [Results and Discussion](#)

[Experiment 2](#)

○ [Method](#)

○ [Results](#)

[Acknowledgments:](#)

[References](#)

Figures/Tables

[Figure 1](#)

[Figure 2](#)

[Figure 3](#)

[Figure 4](#)

[Figure 5](#)

and [Parkin \(1996\)](#) for frequency discrimination and by [Wilson, Finley, Lawson, Wolford, Eddington, and Rabinowitz\(1991\)](#) and [Boex, Pelizzone, and Montandon \(1994\)](#) for speech understanding.

Although Ineraid patients perform better with CIS processors than with analogue processors, they continue to make errors in the identification of consonants and vowels. The aim of Experiment 1 was to determine whether the reports of better consonant recognition with the CIS strategy than with the Ineraid strategy were replicable with a sample of patients who had worn a portable CIS processor for at least a month before testing and, thus, had been allowed to accommodate to the processing scheme. Wilson's patients and the multilingual group tested by [Boex, Pelizzone, and Montandon \(1994\)](#) had used CIS processors for only a few hours in a laboratory setting. The aim of Experiment 2 was to probe the mechanism(s) underlying the errors in consonant recognition.

Experiment 1 [TOP](#)

Method [TOP](#)

Subjects • The subjects were seven patients who had used the Ineraid processor for no less than 4 yr. They ranged in age from 31 to 72 yr.

Stimuli • The stimuli were single exemplars of /b d g p t k s ʃ θ tʃ z m n w l j/ spoken in "aCa" format by a male talker. The stimuli were originally recorded on audio tape and then digitized at a 20 kHz sample rate ([Dorman, Soli, Dankowski, Smith, McCandless, & Parkin, 1990](#)). The stimuli were presented, after D/A conversion and filtering at 9.7 kHz, in blocks of 16 and were randomized within each block. Each block was presented five times. Before the current test administration, the patients had been tested with the consonant set many times through their Ineraid processors and once before with their CIS processor.

Signal processors • The Ineraid prosthesis consists of i) 6 monopolar electrodes implanted in the scala tympani with remote reference, ii) a percutaneous pedestal to which the electrode wires are attached, and iii) a portable speech processing and electrode stimulation system([Eddington, 1980](#)). The most apical electrode is located about 22 mm from the round window. The electrodes are spaced at 4 mm intervals. The four most apical electrodes are activated in most patients. Each of the four electrodes is driven by an analogue signal derived from the input signal after the operation of an AGC circuit and filtering by fourth-order band-pass filters ([Eddington, 1980; Wilson, Finley, Lawson, Wolford, Eddington, & Rabinowitz, 1991](#)). The filter center frequencies are 0.5, 1, 2, and 3.4 kHz. The signal processor has an auxiliary input jack as well as an external microphone. The frequency response of the auxiliary input circuit is similar to that used for the external microphone.

The CIS processor used in this study is an implementation of the [Wilson, Finley, Lawson, Wolford, Eddington, and Rabinowitz\(1991\)](#) processor fabricated at the University of Innsbruck([Zierhofer, Peter, Bril, Pohl, Hochmair-Desoyer, & Hochmair, 1994](#)). The signal processor was a 6-channel design with sixth-order band-pass filters, 400 Hz first-order smoother, and full wave rectification. Channel center frequencies were 393 Hz, 639 Hz, 1037 Hz, 1680 Hz, 2730 Hz, and 4440 Hz. The channels were of equal width on a logarithmic scale. Signals were pre-emphasized above 1200 Hz. Pulse duration and pulse rate were chosen for each patient based on the results of tests of consonant understanding conducted with, most generally, pulse rates of 823, 1170, and 2020 pps and pulse durations ranging from 40 μs/period and 100μs/period. The best combination of pulse rate and pulse duration differed among patients.

Design • On the day before the patients were fitted with a CIS processor, they were tested through their Ineraid processors with a variety of speech materials including the VCV stimuli material described above. At 4 to 6 wk after they had been fitted with a portable CIS processor, they were tested again with the VCV stimuli.

Procedure • The test signals were delivered directly to the patients' signal processors via the auxiliary input jack. The patients were presented two practice lists (with correct answers) before the test sequence began. Responses were collected via a touch-sensitive response pad.

Results and Discussion [TOP](#)

Each patient achieved a higher score with the CIS processor than with the Ineraid processor. The range of improvement was 15 to 49%. The mean percent correct scores, 51% (SD = 9%) for the Ineraid and 81% (SD = 14%) for the CIS processor, differed significantly ($t = 5.49, p < 0.0001$). Thus, our results provide independent confirmation of the [Lawson, Wilson, and Finley \(1993\)](#) and the [Boex, Pelizzone, and Montandon \(1994\)](#) reports of better consonant recognition with the CIS processor.

The group results are shown in confusion-matrix format in [Figure 1](#) for the Ineraid processor and in [Figure 2](#) for the CIS processor. Identification accuracy for each of the manner-by-voicing categories was better with the CIS processor than with the Ineraid processor. Inspection of the results for the CIS processor in [Figure 2](#) reveals that errors of manner and voicing were relatively rare and that errors in place of articulation were the most common. The voiceless stop consonants /ptk/ were, as a group, the most difficult to identify. Individual scores for this consonant group ranged from 40 to 100% correct. The most common error response for /p/ was /t/, the most common error response for /t/ was /k/, and the most common error response for /k/ was /t/.

		R e s p o n s e															
		b	d	g	p	t	k	s	ʃ	θ	tʃ	z	m	n	w	l	j
b		51	3	9	6					17		3	9				3
d		3	63	29		3				3							
g			14	74		3		3		3		3					
p			3	6	6	20	9	6			28	17	6				
t					3	43	51					3					
k					20	21	40			6	3						
s			3			3		28	63		3						
ʃ								9	83		9						
θ		9	3		6			31		43		3			3		
tʃ						9			6		83	3					
z		3	3	3						9	9	65		3	6		
m													49	34	3	6	9
n												3	14	77			6
w												3		20	31	28	17
l													11	11	26	46	6
j				6		6				9			6	6	3	3	62

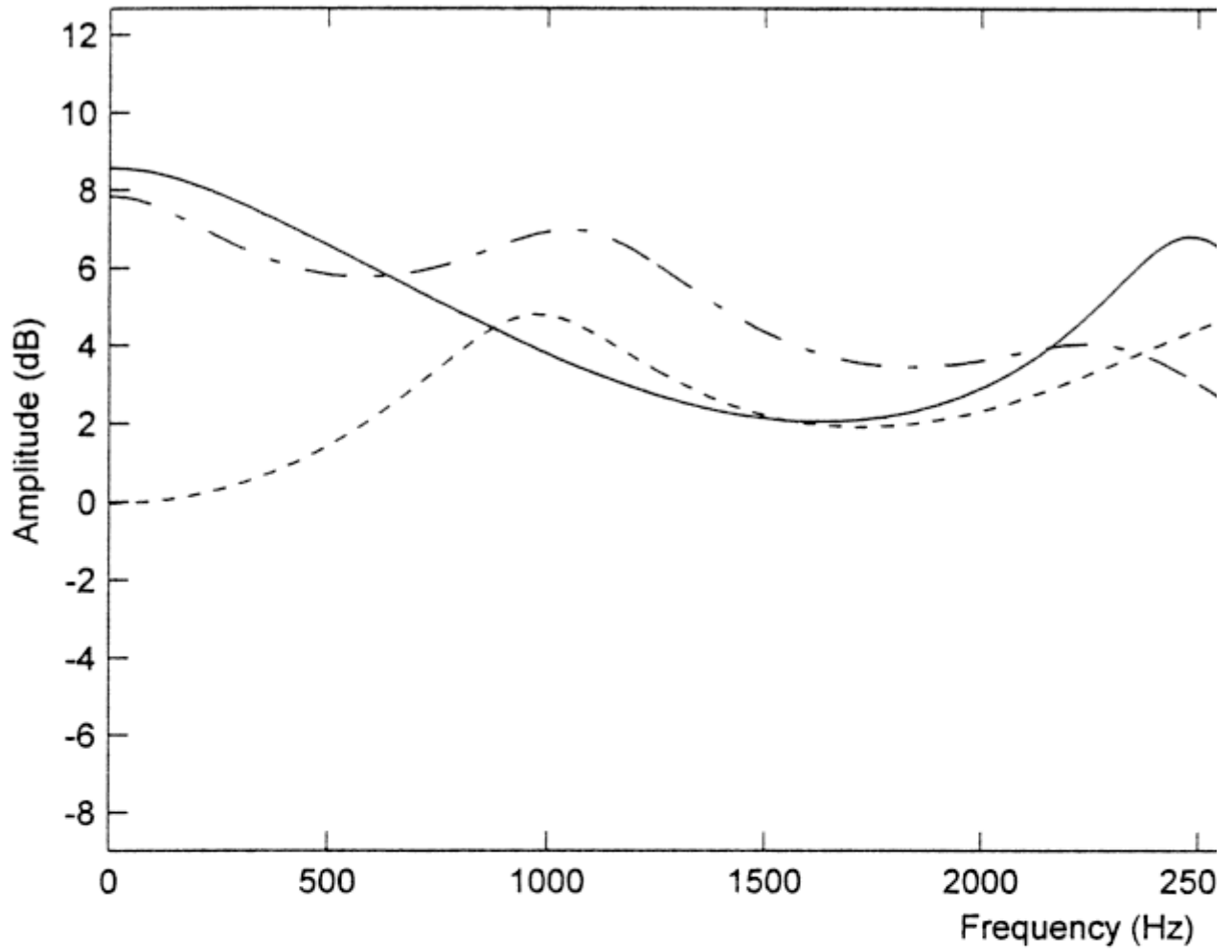
Figure 1. Percent correct scores for consonants presented through the Ineraid processor.

		R e s p o n s e														
		b	d	g	p	t	k	s	ʃ	θ	tʃ	z	m	n	w	l
b	91									9						
d		74	20			3	3									
g			94				6									
p				68	14	3	3	3				6	3			
t				3	57	40										
k				3	26	68										
s						3	86		6		3					
ʃ							9	91								
θ	6			3			9		80			3				
tʃ					11			6			77	6				
z				3								97				
m													77	17	3	3
n									6					94		
w															63	17
l															11	88
j												3	3	6	3	

Figure 2. Percent correct scores for consonants presented through the CIS signal processor.

As a first step toward understanding why the voiceless stops were difficult to identify, /pa/, /ta/, and /ka/ were processed through a software simulation of the CIS processor and the signal level in each of the channels was calculated. The spectra at signal onset for /p/, /t/, and /k/ are shown at the top of [Figure 3](#). Signal levels at stimulus onset as a function of processor channel are shown at the bottom of [Figure 3](#). Inspection of the top and bottom panels indicates that the pattern of activity across the six-electrode array captures the prominent features of the onset spectra in spite of using only six points to define the spectra.

spectra at ons



channel outputs a

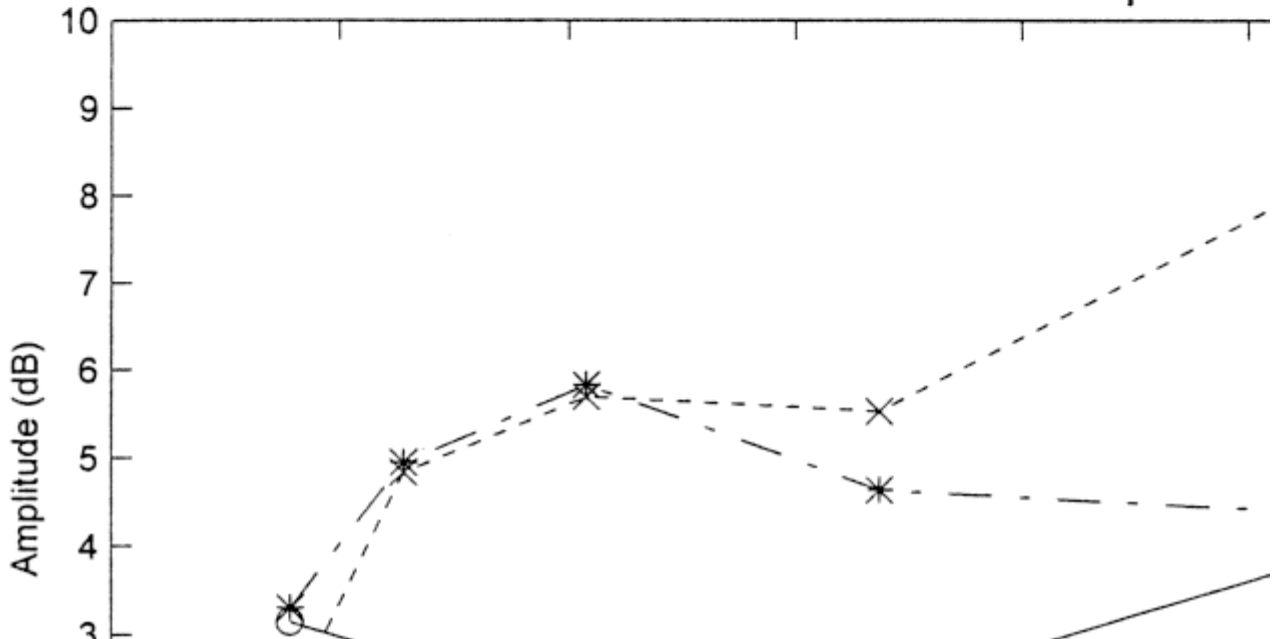


Figure 3. Top: Onset spectra for /p/, /t/, and /k/. The spectra were generated using a 20 msec window starting at the burst onset and a fourteenth-order linear predictive coding analysis. Bottom: Output levels of the CIS processor for /p/, /t/, and /k/. Sequential points along each function represent the output levels of Channels 1 to 6.

As noted above, individual scores for identification of the voiceless stop consonants varied from 40 to 100% correct. This outcome demonstrates that for some patients the differences in output patterns across the six electrode array for /p/, /t/, and /k/ were not large enough for reliable identification of the stimuli. For other patients, e.g., the one with 100% identification accuracy, the differences in output patterns were, by definition, sufficient. On this view, then, patients differ in their ability to use differences in channel outputs in the service of consonant, and, more broadly, phonetic, identification. Given this view, we may be able to increase identification accuracy in a group of patients by accentuating differences in the patterns of channel outputs. To this end, in Experiment 2 the stimuli of Experiment 1 were filtered so as to accentuate the differences in channel output patterns. At issue was whether identification accuracy would improve for the filtered stimuli.

Experiment 2 [TOP](#)

Method [TOP](#)

Subjects • Five patients were tested. They were a subset of the patients tested in Experiment 1 and were chosen for participation solely on the basis of availability for testing.

Stimuli • The tokens of /pa/, /ta/, and /ka/ used in Experiment 1 were used in Experiment 2. In one stimulus set, /pa/, /ta/, and /ka/ were presented in their original form and, in another, were presented after filtering. The filtering operations were based on the following logic. The analysis of data for Experiment 1 indicated that /ta/ was the most common error response for /pa/. It follows that the channel output patterns for /p/ and /t/ must have common features. Inspection of [Figure 3](#) indicates that /p/ and /t/ share an increase in energy from Channel 4 to Channel 5. Reducing the levels of Channels 5 and 6 would make the /p/-spectrum more labial-like, i.e., down-tilted ([Blumstein & Stevens, 1979](#)), and would increase the distinctiveness of /p/ relative to /t/. The analysis of data from Experiment 1 also indicated that /ka/ was the most common error response for /ta/ and that /ta/ was the most common error response for /ka/. Inspection of [Figure 3](#) reveals that the token of /k/ has a relatively high level of energy in Channel 6, which increases its similarity to the pattern of energy for /t/. Reducing the energy in Channels 5 and 6 of /k/ would emphasize the "mid-frequency" peak characteristic of velars ([Blumstein & Stevens, 1979](#)) and would increase the difference between the patterns of stimulation for /k/ and /t/.

The stimulus manipulations described above were accomplished by digital filtering of the tokens of /pa/, /ta/, and /ka/ used in Experiment 1. The results of the filtering operations are shown in [Figure 4](#). Inspection of the top panel in [Figure 4](#) reveals that the filtered version of /p/ had reduced energy in Channels 4, 5, and 6. Inspection of the middle panel in [Figure 4](#) indicates that the filtered version of /t/ had reduced energy in Channels 1, 2, 3, and 4. Inspection of the bottom panel in [Figure 4](#) indicates that the filtered version of /k/ had reduced energy in Channels 5 and 6. As a result, the output patterns for the filtered stimuli were more dissimilar than were the output patterns for the unfiltered stimuli.

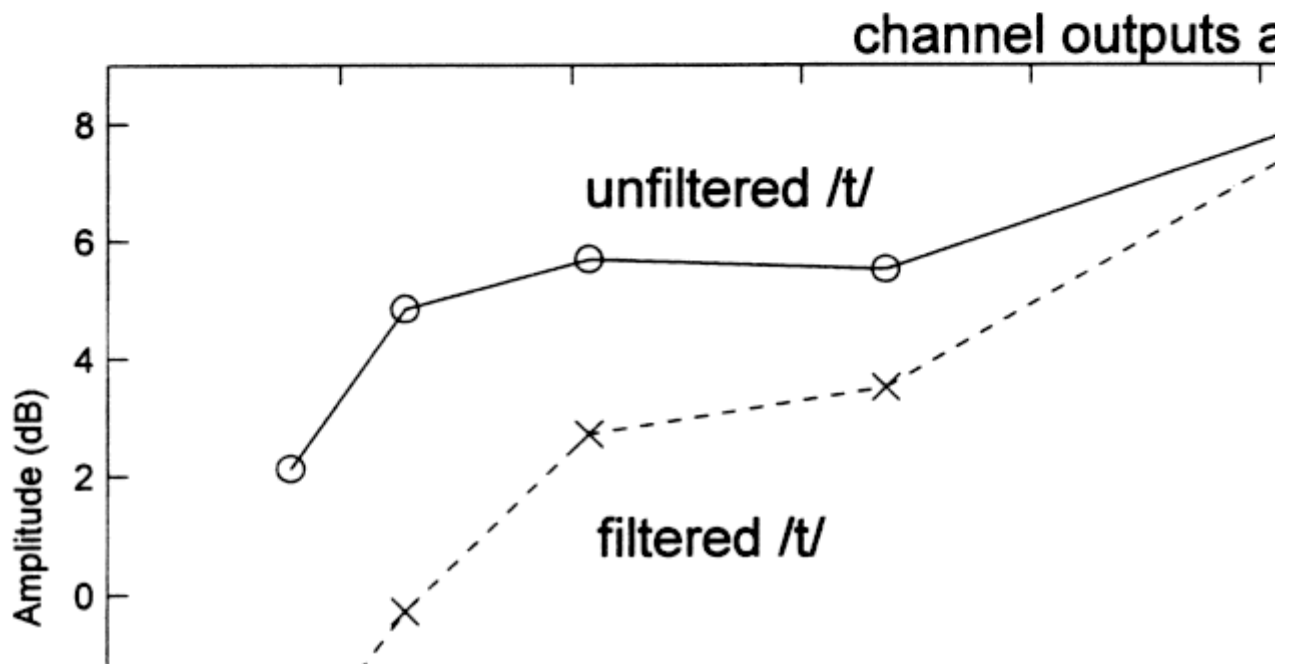
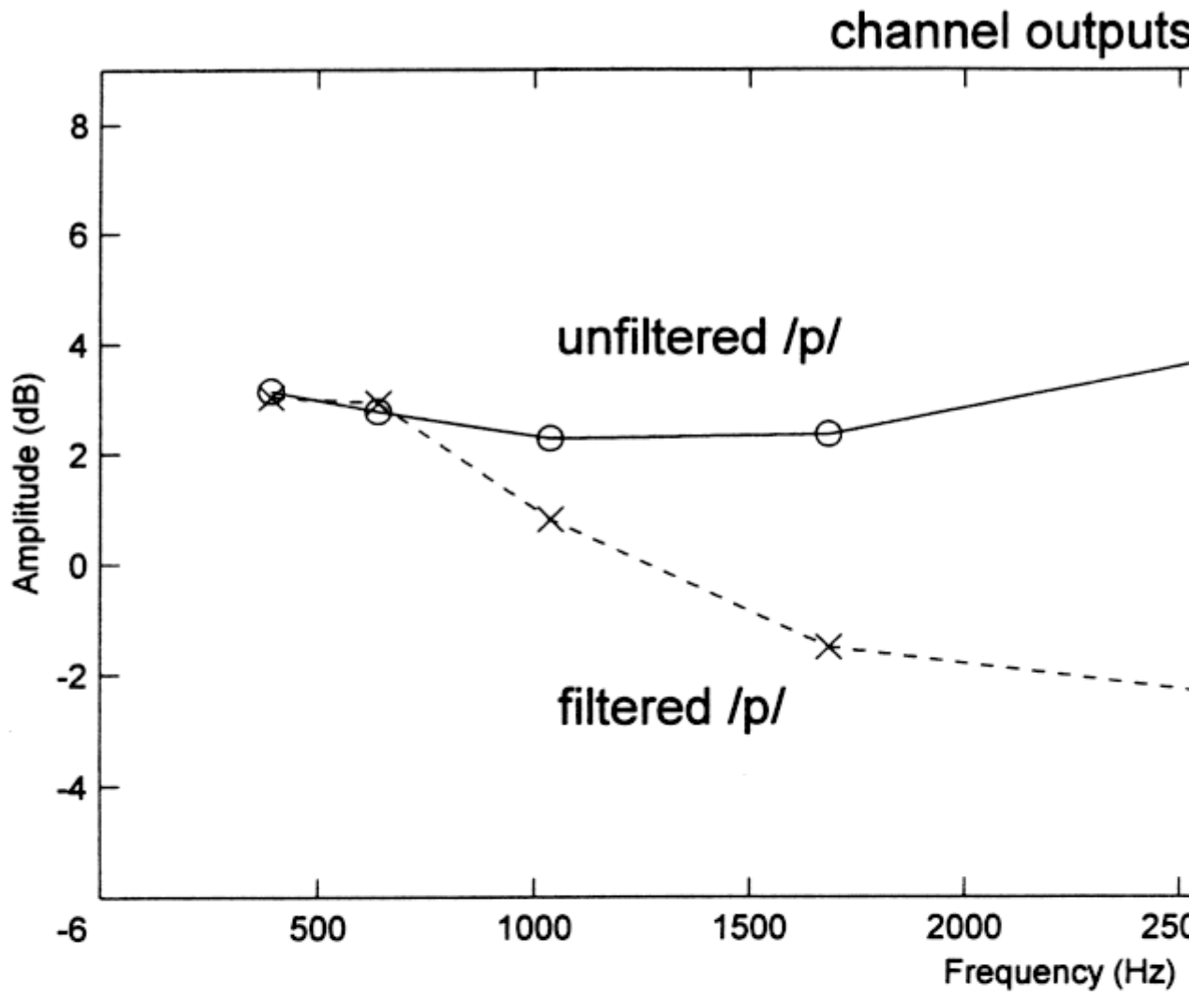


Figure 4. Output levels for channels 1 to 6 of the CIS processor for the normal, i.e., unfiltered, and the filtered version of /p/ (top), for the normal and the filtered version of /t/ (middle), and for the normal and the filtered version of /k/ (bottom).

Procedure • The unfiltered and filtered versions of /pa/, /ta/, and /ka/ were combined into a single test sequence. Each stimulus was presented five times. The stimuli were completely randomized in the test sequence. Because the results of Experiment 1 indicated that more than 90% of the error responses were errors of place, and not manner or voicing, the patients were told that the stimuli would be versions of /pa/, /ta/, and /ka/ and were instructed to respond only "p," "t," or "k." The test signals were delivered directly to the patients' signal processors via the auxiliary input jack. Responses were collected via a touch-sensitive response pad.

Results ^{TOP}

The results are shown in [Figure 5](#), top and bottom. The mean score for the unfiltered stimuli was 66% (SD = 13%) correct. This score is similar to the 64% correct found for voiceless stop consonants in Experiment 1 and indicates that the restriction in the number of possible responses did not significantly alter the experimental outcome. The mean score for the filtered stimuli was 80% (SD = 16%) correct. Four of the five patients achieved a higher score in the filtered condition than in the unfiltered condition. The range of improvement was 5 to 27%. One patient achieved the same score in both conditions. The mean score for the filtered condition (80%) was significantly higher than the mean score (64%) for the unfiltered condition ($t = -2.82, p = 0.02$).

	Response		
	p	t	k
unfiltered			
p	76	20	4
t		60	40
k		44	56
filtered			
p	96	4	4
t	16	76	8
k	4	20	76

Figure 5. Percent correct scores for unfiltered and filtered (enhanced) /pa/, /ta/, and /ka/.

Discussion • Implant patients who use multichannel devices receive information about consonant identity in two forms. One is the time/amplitude envelope. The other is the spatial distribution of energy along the electrically stimulated cochlea. If patients fail to identify consonants within a manner category, e.g., the place of articulation of voiceless stops, it is likely that the patients are not able to use the normally occurring differences in spatial distribution of energy along the cochlea in the service of identification. We make this assumption because amplitude envelope information plays more of a role in the identification of manner and voicing than in the identification of place of articulation (Van Tassel, Soli, Kirby, & Widin, 1987). To assess whether this assumption is true, in Experiment 2 we enhanced the differences in channel output levels for three phones (/ptk/) which, in Experiment 1, were found to be difficult to identify. We found that patients achieved a higher level of identification accuracy for the stimuli with enhanced differences in channel output levels. Thus, we confirm that

one factor underlying poor identification of consonants is an inability to use normally occurring differences in channel output levels as indicators of consonant identity. On this logic, individual differences in consonant identification accuracy are based on individual differences in the ability to discriminate among channel output levels.

The filtering operations we have used to improve consonant recognition are not likely to be implemented in a wearable signal processor. The limitation lies not in the filtering operations per se but in the accuracy of the steps of segmentation of the utterance and recognition of phonetic segments that would precede the filtering operations. Given that this is the case, what have we accomplished by demonstrating that we can improve the recognition of a small set of consonants by filtering? First, we have shown that the spatial pattern of energy along the electrically stimulated cochlea encodes information about consonant place of articulation. Second, we have shown that a small number of channels would be sufficient to support very high levels of recognition of consonant place of articulation if a signal processor's channel outputs could be made more distinct. Finally, we have shown that we have not reached the limit of speech understanding that can be supported by the population of neural elements remaining in our patients' auditory systems. Patients, especially those who use the Ineraid, can be told that the future holds the promise of better speech understanding.

Acknowledgments: [TOP](#)

This research was supported by NIDCD DC00654-05.

References [TOP](#)

Blumstein, S., & Stevens, K. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, 66, 1001-1017.

[\[Context Link\]](#)

Boex, C., Pelizzone, M., & Montandon, P. (1994). Improvements in speech recognition with the CIS strategy for the Ineraid multichannel intracochlear implant. In I. Hochmair-Desoyer & E. Hochmair (Eds.), *Advances in Cochlear Implants* (pp. 136-140). Vienna: International Interscience Seminars.

[\[Context Link\]](#)

Dorman, M. (1994). Mechanisms of formant frequency resolution with the Ineraid cochlear implant. In I. Hochmair-Desoyer & E. Hochmair (Eds.), *Advances in Cochlear Implants* (pp. 302-305). Mantz: International Interscience Seminars.

[\[Context Link\]](#)

Dorman, M., Smith, L., Smith, M., & Parkin, J. (1996). Frequency discrimination, and speech recognition by patients who use the Ineraid, and CIS cochlear-implant signal processors. *Journal of the Acoustical Society of America*, 99, 1174-1184.

[\[Context Link\]](#)

Dorman, M., Soli, S., Dankowski, K., Smith, L., McCandless, G., & Parkin, J. (1990). Acoustic cues for consonant identification by patients who use the Ineraid cochlear implant. *Journal of the Acoustical Society of America*, 88, 2074-2079.

[\[Context Link\]](#)

Eddington, D. (1980). Speech discrimination in deaf subjects with cochlear implants. *Journal of the Acoustical Society of America*, 68, 885-891.

[\[Context Link\]](#)

Lawson, D., Wilson, B., & Finley, C. (1993). New processing strategies for multichannel cochlear prostheses. In J. Allum, D. Allum-Mecklenburg, F. Harris, & R. Probst, (Eds.), *Progress in Brain Research*, Vol. 97 (pp. 313-321). New York: Elsevier Science Publishers.

[\[Context Link\]](#)

Van Tassel, D., Soli, S., Kirby, V., & Widin, G. (1987). Speech waveform envelope cues for consonant recognition. *Journal of the Acoustical Society of America*, 82, 1152-1161.

[\[Context Link\]](#)

Wilson, B., Finley, C., Lawson, D., Wolford, R., Eddington, D., & Rabinowitz, W. (1991). Better speech recognition with cochlear implants. *Nature*, 352, 236-238.

[\[Medline Link\]](#) [\[CrossRef\]](#) [\[Context Link\]](#)

Zierhofer, C., Peter, O., Brill, S., Pohl, P., Hochmair-Desoyer, I., & Hochmair, E. (1994). A multichannel cochlear implant system for high-rate pulsatile stimulation strategies. In I. Hochmair-Desoyer & E. Hochmair (Eds.), *Advances in Cochlear Implants* (pp. 204-207). Mantz: International Interscience Seminars.

[\[Context Link\]](#)

