

The coding of vowel identity by patients who use the Ineraid cochlear implant

Michael F. Dorman

Arizona State University, Tempe, Arizona 85287-0102 and University of Utah School of Medicine, Salt Lake City, Utah 84132

Luther Smith

Sinai Samaritan Medical Center, Milwaukee, Wisconsin 53233

Michael Smith

Arizona State University, Tempe, Arizona 85287

James Parkin

University of Utah School of Medicine, Salt Lake City, Utah 84132

(Received 13 February 1992; revised 3 July 1992; accepted 13 August 1992)

The aim of this experiment was to determine whether patients who use the Ineraid cochlear implant code formant frequency by means of the information in the time waveforms presented to the electrodes, or by the information in the distribution of energy among the electrodes. Six patients were presented three vowels for identification. The vowels were presented in two configurations. In one, the temporal cues and distribution of energy cues were both appropriate for a given vowel. In the other, the temporal information specified one vowel while the distribution of energy specified another vowel. The results of the identification task indicated that patients relied on information contained in the distribution of energy for vowel identification. These results can be interpreted as indicating that the Ineraid overcomes the limited spatial resolution afforded by four monopolar electrodes by using the balance of energy among electrodes to produce a continuous, rather than discrete, coding of cochlear place of stimulation.

PACS numbers: 43.71.Ky, 43.71.Cq, 43.66.Ts, 43.64.Me

INTRODUCTION

The aim of this experiment was to assess how patients who use the Ineraid cochlear implant code formant frequency in the service of vowel recognition.

Frequency coding and frequency resolution could be a function of place coding, temporal coding or a combination of both. When a signal arrives at the Ineraid's signal processor, the signal is filtered into bands with center frequencies of 0.5, 1.0, 2.0, and 3.4 kHz (Eddington, 1980). The filter skirts are broad—6 dB/oct. Analog representations of the waveforms in the four bands are then delivered to electrodes 1–4 which are arrayed from apex to base in the cochlea. The most apical electrode is usually 20–22 mm from the round window and the electrodes are spaced at 4-mm intervals.

The frequency of the first formant of a speech signal (F_1) could be specified by the time waveform in the most apical channel. As shown in Fig. 1, the representation of F_1 for signals with F_1 s of 350, 500, and 700 Hz is well preserved in the time waveforms presented to electrode 1. However, the more dominant feature of the time waveform is the representation of the pitch period, or F_0 . In order for temporal encoding of F_1 to occur, some neural elements would have to

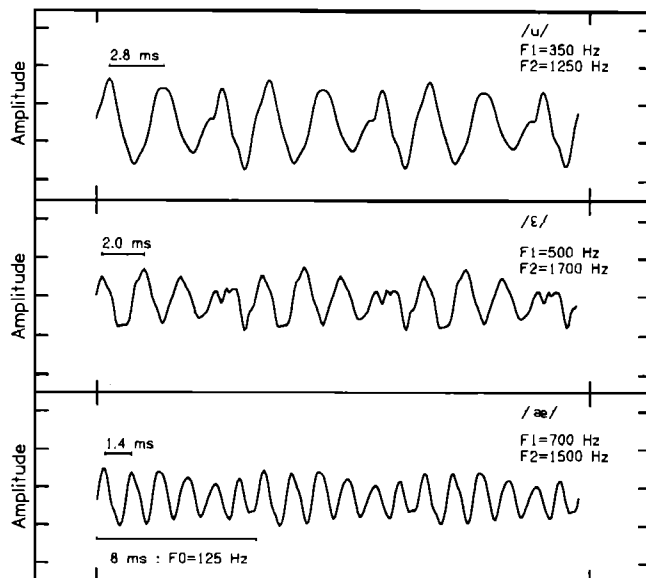


FIG. 1. Time waveforms from channel 1 of the Ineraid for a three pitch-period segment from three vowels. The period of F_1 is indicated at the top of each waveform. The period of F_0 is indicated below the waveform in the lowest panel.

synchronize to the period of $F1$ while others synchronized to the period of $F0$.

Temporal coding of $F2$ and higher frequencies is unlikely. The coding of temporal intervals by a single fiber is limited by the neural refractory period of about 1 ms. However, groups of fibers firing in parallel may be able to code temporal intervals less than 1 ms in duration (Hartmann and Klinke, 1990).

The possibility that $F1$ can be derived from a temporal code is confirmed by the ability of patients who use the Vienna single-channel implant to identify vowels. Von Wallenberg *et al.* (1985) and Von Wallenberg *et al.* (1990) describe patients whose patterns of vowel identification reflect resolution of $F1$'s at approximately 300, 500, and 750 Hz. Rosen and Ball (1986) and Tyler *et al.* (1989b) report broadly similar findings. Some Vienna patients also appear to make use of information from $F2$. It is not clear how $F2$ information is encoded by these patients.

For Ineraid patients, the frequency of $F1$ might also be specified by a place code. Consider the effect of a $F1$ at 300 Hz presented to the Ineraid processor. The output of the processor would have a high rms level in channel 1 and lower levels in channels 2–4. As the frequency of $F1$ increases from 300 to 500 or 700 Hz, the rms level in channel 1 would increase slightly and then decrease slightly, depending on the filter shape and the formant levels, but the energy in channel 2 would increase. *Thus, the frequency of $F1$ could be specified by the balance of energy in channels 1 and 2.* That is, as $F1$ frequency increased, the balance of energy would tip toward electrode 2. In similar fashion, the frequency of $F2/F3$ could be specified by the balance of energy among channels 2–4.

There is some evidence to support the suggestion that frequency can be coded by the distribution of energy among channels of a multichannel implant. Townshend *et al.* (1987) have reported, for two patients, that when two electrodes are stimulated simultaneously, and the balance of current is varied between the two electrodes, then pitch varies from higher to lower as the current balance swings from the more basal electrode to the more apical electrode.

The aim of the present experiment was to assess whether Ineraid patients code formant frequency by the information in the time waveforms presented to the electrodes, or by the information contained in the balance of energy among electrodes. We have focused on differences in $F1$ among vowels because the data from single-channel implant patients demonstrates that a temporal code can be used for recognition of $F1$ and because analyses of vowel recognition by Ineraid patients indicates that differences in $F1$ are used in recognition (see Eddington, 1983, and later studies by Dorman *et al.*, 1989; Tyler *et al.*, 1989a,c; and Wilson *et al.*, 1990).

We constructed three vowels in two configurations. The vowels /u ε æ/ were chosen so that the frequency of $F1$ would be maximally different among the signals ($F1 = 350, 500, \text{ and } 700 \text{ Hz}$, respectively) and, thus, so that the time waveforms in channel 1 would be maximally different. In one test configuration, both the temporal information in each channel and the distribution of energy among the channels specified the same vowel. In the other condition—a conflicting cue condition—we altered the amplitude of the sig-

nals in channels 1–4 so that the temporal waveform in each channel specified one vowel, for example /ε/, but the distribution of energy in the channels specified another vowel, for example, /u/ or /æ/. At issue was whether our patients would identify the vowels on the basis of the time waveforms in the channels or would identify the vowels on the basis of the distribution of rms energy among the channels.

I. METHOD

A. Subjects

The subjects were six individuals who used the Ineraid prosthesis. The subjects were chosen from a pool of 20 patients on the basis of their vowel identification ability. We wished to test patients who could identify the vowels /u/, /ε/, /æ/ with near perfect accuracy. All of the patients had previously participated in psychoacoustic tests and tests of speech recognition.

B. Stimuli

The stimuli were three steady-state vowels—/u/, /ε/, and /æ/. The vowels were synthesized in cascade mode using a version of the KLATT synthesizer. The signals were created with identical rise-fall time (30 ms), duration (200 ms) and pitch contour. The overall signal amplitude varied in normal fashion. The formant frequencies for the three vowels are shown in Table I.

To determine the rms level by channel for each vowel, the prototype vowels were directed to the Ineraid via the auxiliary input jack. The Ineraid's four outputs were detected across series 1-kΩ resistors, amplified and digitized (Data Translation DT2801-A). The signals were displayed via a waveform editing program (WAVED; Boys Town National Hospital). With the aid of the program, rms levels were assessed in each of the four channels for the three prototype vowels. The relative levels are shown in Fig. 2.

C. Stimulus modification

In order to create signals with conflicting cues to vowel identity, four computer-controlled digital attenuators (Turner-Davis Technologies) were inserted in the circuit between the Ineraid processor and the patient. Signals were first filtered into four bands by the Ineraid. Next, the signals in the four bands were directed to four preamplifiers and to the four digital attenuators. The signals were then passed to four optical isolation amplifiers for patient isolation and safety. Finally, the signals were delivered to the patients' electrodes. This arrangement allowed the experimenter to

TABLE I. Formant frequencies for the three prototype vowels.

	$F1$	$F2$	$F3$
/u/	350	1250	2200
/ε/	500	1700	2500
/æ/	700	1500	2400

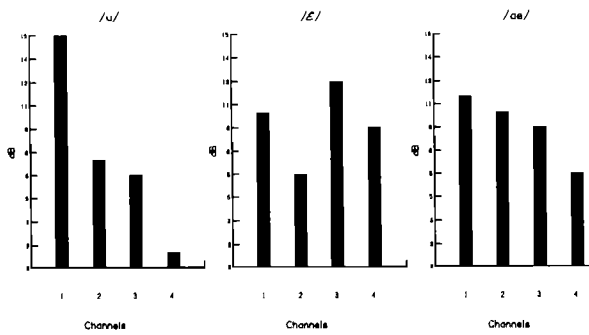


FIG. 2. rms levels in each channel of the Ineraid for three vowels.

use the digital attenuators to vary the signal level on a given electrode over a ± 10 -dB range. This range was sufficient to give a prototype vowel the amplitude characteristics of each of the other vowels.

In the test sequence, each vowel was presented in three amplitude configurations. For example, for / ϵ /, one configuration was the unaltered prototype. In another configuration, / ϵ / was given the amplitude profile of / u / (in this case / u / was the "target vowel"). In the third configuration, / ϵ / was given the amplitude profile of / æ / (in this case / æ / was the target vowel). Similar manipulations were performed with / u / and / æ /. Five repetitions of each stimulus were generated for the identification test.

D. Procedure

The patients were first presented five repetitions of each of the prototype vowels in a practice sequence. The identity of each stimulus was indicated to the patient. Next, the patients were presented the test stimuli, including the prototypes, for identification in a randomized sequence. The patients indicated the identity of each stimulus on a touch sensitive pad. The pad contained a "repeat" key and each stimulus could be heard as many times as the patient wished. The response choices were / u /, / æ /, / ϵ /.

II. RESULTS

The mean identification score for each stimulus is shown in Table II. Two types of data analyses were performed. The first assessed whether the number of prototype responses was reduced when the amplitude profile of the prototype vowel was altered. As shown in Table II the prototype vowels were identified with near perfect accuracy. In five of the six conflicting-cue conditions *all* patients evidenced a reduction in the number of prototype responses. In the case of / æ / given the amplitude profile of / ϵ /, four of six patients evidenced a reduction in the number of prototype responses. Thus changing the amplitude profile of the prototype vowels significantly reduced identification accuracy.

The second data analysis assessed whether a change in amplitude profile elicited a particular response or merely made the prototype vowel difficult to identify. To answer this question the number of target vowel responses were compared to the sum of nontarget vowel responses. For example, in the case of / u / given the amplitude profile of / ϵ /

TABLE II. Percent response as a function of amplitude profile. For signals with entries of the form / \rightarrow /, the second member of the expression indicates the vowel whose amplitude profile was mimicked.

Stimuli	Response		
	u	ϵ	æ
/u/	93	3	3
/ ϵ /	0	97	3
/ æ /	0	3	97
/u/ \rightarrow / ϵ /	0	84	16
/u/ \rightarrow / æ /	0	63	37
/ ϵ / \rightarrow /u/	93	3	3
/ ϵ / \rightarrow / æ /	10	23	67
/ æ / \rightarrow /u/	90	7	3
/ æ / \rightarrow / ϵ /	0	47	53

(/u/ \rightarrow / ϵ /), the number of /u/ and / æ / responses were combined to produce the nontarget response category. In four of the six conditions, the target vowel was the most probable response (/u/ \rightarrow / ϵ /; $t = 3.7$, $p = 0.006$; / ϵ / \rightarrow /u/; $t = 6.5$, $p = 0.006$; / ϵ / \rightarrow / æ /; $t = 2.33$, $p = 0.033$; / æ / \rightarrow /u/; $t = 8.94$, $p = 0.0001$).

In two cases, /u/ \rightarrow / æ / and / æ / \rightarrow / ϵ /, the target vowel did not elicit more responses than nontarget vowels ($t = -0.948$, $p = 0.806$ and $t = -2.23$, $p = 0.58$, respectively). However, in the case of / æ / \rightarrow / ϵ /, half of the patients (three of six) did give the target vowel as the majority response.

III. DISCUSSION

Our aim in this experiment was to determine whether patients who use the four-channel Ineraid implant base vowel identification on the temporal information about formant frequency in the channels, or base identification on the distribution of energy among the channels. We have found that when the amplitude profile of one vowel is changed to that of a second vowel, leaving the temporal cues unchanged, then, most generally, patients report the second vowel. This outcome suggests that the pattern of energy along the electrically stimulated cochlea is the principle cue to vowel identity.

In two of six instances, the distribution of energy for a signal did not predict the majority response. In one instance, /u/ with the amplitude profile of / æ /, for which the majority response was / ϵ /, it would be possible to speculate that temporal coding of the 350-Hz first formant of /u/, and the distribution of energy cue to / æ /, might have led to a hybrid response of / ϵ /, i.e., a vowel with a F_1 intermediate between that of /u/ and / æ /. However, such a feature-sharing account cannot be offered for the failure of the / æ / \rightarrow / ϵ / manipulation to produce a majority of / ϵ / responses. In this instance, we suppose that, for reasons we do not clearly understand, we did not reproduce the distribution of energy for / ϵ / appropriately for half (three of six) of the patients.

A reviewer has noted a possibility which would weaken our conclusion that temporal factors play little role in the coding of vowel formant frequencies. That possibility is that

when we altered the channel amplitudes, the change in current interactions within the cochlea altered the temporal waveforms. We have no way of knowing whether this is a reasonable possibility or not. Consider, however, the /ε/ → /u/ manipulation in which all patients reported /u/. In this case, the amplitude of channel 1 was increased while that of channel 2 was left about the same. It is difficult to imagine an interaction which would produce a poorer representation of the temporal information in channel 1.

ACKNOWLEDGMENT

This research was supported by Grant No. DC-000654 from the National Institute on Deafness and Other Communicative Disorders.

- Dorman, M., Dankowski, K., McCandless, G., and Smith, L. (1989). "Identification of synthetic vowels by patients using the Symbion multi-channel cochlear implant," *Ear Hear.* **10**, 40-43.
- Eddington D. (1980). "Speech discrimination in deaf subjects with cochlear implants," *J. Acoust. Soc. Am.* **68**, 885-891.
- Eddington, D. (1983). "Speech recognition in deaf subjects with multi-channel intracochlear electrodes," in *Cochlear Prostheses: An International Symposium*, edited by C. Parkins and S. Anderson (NYAS, New York).
- Hartmann, R., and Klinke, R. (1990). "Response characteristics of nerve fibers to patterned electrical stimulation," in *Cochlear Implants: Models of the Electrically Stimulated Ear*, edited by J. Miller and F. Spelman (Springer-Verlag, New York).
- Rosen, S., and Ball, V. (1986). "Speech perception with the Vienna extra-cochlear single-channel implant: A comparison of two approaches to speech coding," *Brit. J. Audiol.* **20**, 61-83.
- Townshend, B., Cotter, N., Compennolle, D., and White, R. (1987). "Pitch perception by cochlear implant subjects," *J. Acoust. Soc. Am.* **82**, 106-115.
- Tyler, R., Moore, B., and Kuk, F. (1989a). "Performance of some of the better cochlear-implant patients," *J. Speech Hear. Res.* **32**, 887-911.
- Tyler, R., Tye-Murray, N., Moore, B., and McCabe, B. (1989b). "Synthetic two-formant vowel perception by some of the better cochlear-implant patients," *Audiology* **28**, 301-315.
- Tyler, R., Tye-Murray, N., and Otto, S. (1989c). "The recognition of vowels differing by a single formant by cochlear-implant subjects," *J. Acoust. Soc. Am.* **86**, 2107-2112.
- von Wallenberg, E., Hochmair-Desoyer, I., and Hochmair, E. (1985). "Speech processing for cochlear implants," in *Proceedings of the Seventh Annual Conference on the IEEE Engineering in Medicine and Biology Society*, pp. 1114-1118.
- von Wallenberg, E., Hochmair, E., and Hochmair-Desoyer, I. (1990). "Initial results with simultaneous analogue and pulsatile stimulation of the cochlea," *Acta Oto. Suppl.* **469**, 140-149.
- Wilson, B., Lawson, D., and Finley, C. (1990). "Speech processors for auditory prostheses," Fourth Q. Progr. Rep. on NIH Project N01-DC-9-2401, pp. 4-24.