THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 1991, *43A* (3) 585–601

Vowel and Consonant Recognition with the Aid of a Multichannel Cochlear Implant

Michael F. Dorman
Arizona State University, Tempe, Arizona, U.S.A.

Korine Dankowski
Richards Medical, Memphis, Tennessee, U.S.A.

Geary McCandless, James L. Parkin
University of Utah School of Medicine, Salt Lake City, Utah, U.S.A.

Luther Smith
Sinai Samaritan Medical Center Milwaukee, Wisconsin, U.S.A.

In this report we review the vowel and consonant recognition ability of patients who use a multichannel cochlear implant and who achieve relatively good word identification scores. The results suggest that vowel recognition is accomplished by good resolution of the frequency of the first formant (F1) combined with poor resolution of the frequency of the second formant (F2). The results also suggest that consonant recognition is accomplished (1) by using information from the amplitude envelope, including periodicity/aperiodicity, as cues to manner and voicing, (2) by using F1 as an aid to the identification of manner and voicing, and (3) by using information from cochlear place of stimulation to provide a very crude indication of the shape of the frequency spectrum above 1 kHz.

In this report we describe the word, consonant, and vowel recognition ability of patients who use the Ineraid multichannel cochlear implant. The patients were individuals who had lost all functional hearing bilaterally following insult to the peripheral auditory system. The implant was designed to stimulate electrically elements of the auditory nerve and, in so doing, to

---

Requests for reprints should be sent to: Prof. M.F. Dorman, Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287–0102 U.S.A.

Support for this research was provided by the National Institute for Deafness (RO1–00654).

restore to the patients a measure of auditory sensitivity and some degree of speech and non-speech recognition. Our aim in this report is to infer from the results of several experiments the information in the speech signal that is used by these patients to identify the elementary units of speech, that is, the consonants and vowels. In the first part, we review previous research. In the second, we present the results of two new experiments.

Implant Design

The Ineraid prosthesis consists of (1) 6 monopolar electrodes implanted in the scala tympani with remote reference; (2) a percutaneous pedestal to which the electrode wires are attached, and (3) 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 electrode is driven by an analogue signal derived from the input signal after the operation of an automatic gain control (AGC) circuit and bandpass filtering. The AGC circuit has a very rapid attack time and a relatively slow release time. The centre frequencies of the filters for channels 1–4 (most-apical to most-basal electrodes) are 0.5 1, 2, and 3.4 kHz. The filters roll off at 6 dB/octave.

Given the depth of insertion of the electrodes and their spacing, the bandpassed input to each electrode may be a half octave or more lower in frequency than the corresponding place of cochlear stimulation (see Greenwood, 1990, for a discussion of cochlear frequency position functions).

Coding of Information

Information about consonant and vowel identity is present in the time, amplitude, and frequency domains of an acoustic signal. Shannon (1986) demonstrated in a variety of tasks that temporal resolution can be similar for implant patients and for normal hearing listeners when signals are presented at similar loudness levels. However, Hochmair-Desoyer, Hochmair, and Stiglbrunner (1985) and Tyler, Moore and Kuk (1989a) have found large individual differences among implant patients when temporal resolution is assessed by gap detection.

All implant patients evidence a very reduced dynamic range for signal amplitude (dynamic range is the difference in signal levels sufficient to elicit a response at threshold and to elicit a response of uncomfortable loudness). Hochmair-Desoyer et al. (1985) report dynamic ranges as small as 3 dB and only as large as 25 dB. Dankowski and McCandless (1990) report dynamic ranges between 6 and 17 dB for patients who use the Ineraid.

Implant patients extract only a limited amount of information about stimulus frequency from the acoustic signal. For implant patients, frequency is specified by "rate" and/or "place" codes.

Rate coding is accomplished by auditory fibers discharging in synchrony, or phase locking, to the period of an electrically delivered stimulus. Significant phase locking has been reported for frequencies greater than 1 kHz (Dynes, 1989; Hartmann & Klinke, 1990). However, psychophysical studies with implant patients who use a single electrode prosthesis indicate that pitch increases with pulse rate to only approximately 1000 pulses per sec (pps). For some patients, the perception of pitch may asymptote at a rate as low as 175 pps (Hochmair-Desoyer, Hochmair, Burian, & Stiglbrunner, 1983; Townshend, Cotter, Van Campernolle, & White, 1987). Difference limens for pitch can be very small (10%) for pulse rates up to 300 pps but are much larger (15–40%) for pulse rates near 1000 pps.

The spectral properties of a signal can also be encoded along a scale of relative frequency by information about the *place* of cochlear stimulation. Eddington (1980) assessed the relative pitch of signals presented to a series of monopolar electrodes arrayed from basal to apical locations in the scala tympani. For a given repetition rate, relative pitch increased in a non-linear fashion as more basal electrodes were stimulated. Thus, if signals are bandpass filtered, as in the Ineraid prosthesis, and the energy in the bands is directed to an array of intracochlear electrodes, the rank ordering of the frequency bands can be preserved.

For many Ineraid patients, the number of pitch percepts at frequencies greater than 1 kHz is not limited to the number of channels in the prosthesis or electrodes in the cochlea. For example, when signals are presented in 300-Hz increments between 1 and 2 kHz, patients report hearing changes in pitch. We may account for this in the following way: Recall that signals are delivered to the electrodes after bandpass filtering at 6 dB/octave. The shallow skirts of the filters ensure that a given signal will present energy to more than one electrode. Thus, as signals are varied between 1 kHz and 2 kHz, the distribution of current at electrodes 1–4 will vary. The changes in the distribution of current along the cochlea may be the mechanism that underlies the changes in pitch (see Townshend et al., 1987).

Patients who use the Ineraid prosthesis appear to use a combination of rate and place coding for frequency analysis. Smith (1988) reported that when pure tone signals were directed to a single electrode, pitch estimates reached asymptote at stimulation rates between 400 Hz and 1 kHz. When pure tones were directed through the signal processing system of the Ineraid to all four electrodes, pitch estimates increased with stimulation rate up to 2–3 kHz. Presumably, the increase in range of pitch estimates resulted from place coding of frequencies higher than 1 kHz.

Word Recognition

Patients who use the Ineraid prosthesis have been tested with a variety of word lists. Dorman, Dankowski, Smith, and McCandless (1989b) reported scores ranging from 0 to 100% correct for 50 patients tested with two-syllable (or spondee) words. The median score was 44% correct. Dorman et al. also reported scores ranging from 0 to 60% correct for a more difficult single-syllable word test (the NU-6 word list). The median score on this test was 14% correct. The relatively high scores posted by the best patients suggests that a significant amount of information about consonant and vowel identity can be extracted from the auditory representation of the speech signal (see also Gantz et al., 1988, and Tyler et al., 1989a, for discussions of word recognition by implant patients).

Recognition of Vowels

For normal hearing listeners, the principal cues to vowel identity are the frequencies of the first, second, and third formants of the speech signal. These formant frequencies cover a range of approximately 3.0 kHz. For the "front" vowels in English, the vowels in "beet", "bit", "bait", "bet", "bat", good resolution of low-frequency information (under 600 Hz) combined with poor resolution of higher-frequency information is sufficient for recognition (Dubno & Dorman, 1987). In the extreme condition of having *only* F1 present in the signal, some vowels can still be identified with accuracy greater than chance by some listeners. Dorman and Dubno reported that two of ten normal hearing patients scored greater than 90% correct for a set of five front vowels created with only F1.

The data reviewed above suggest that vowel identification scores for multichannel implant patients would be better than chance if F1 frequency were well resolved and some information about F2 frequency were available.

We have conducted several experiments to assess the ability of implantees to recognize vowels. Our first experiment was conducted with a single subject who had achieved high scores on tests of word recognition (Dorman et al., 1988). The stimuli were 12 computer-synthesized words that shared a common initial and final consonant (/b/ and /t/, respectively), and which differed in the medial vowel (e.g. "beet", "bit", "bet"). Three sets of the 12 stimuli were created. For one, all five formants were synthesized. For a second, only the first formant was synthesized. For a third, only F2–F5 were synthesized. Vowel duration was constant for all sets. The patient was 79% correct for the signals with five formants, 33% correct for the signals synthesized with only F1, and 41% correct for the signals synthesized with only F2–F5. These results indicate that information about both the first and higher formants can be available for use by patients in the task of vowel identification (see also Tyler, Tye-Murray, & Otto, 1989b, for an elegant demonstration of the use of the first and second formants in vowel identification).

In a second experiment, eight patients who had achieved scores of greater than 70% correct on a test of word recognition were presented the 12 synthesized words for identification (Dorman, Dankowski, Smith, & McCandless, 1989a). The combined confusion matrix for the patients is shown in Figure 1. The range of scores was 49 to 79% correct. The mean score was 60% correct. The vowels in "beet", "bit", "bait", "bite", "bet", "bat", "Bert", "boot" were relatively well identified (mean score = 70% correct). The vowels in "but", "bought", "boat" were not well identified (mean score = 29% correct). These vowels were characterized by relatively low-frequency F2s (858–1200 Hz) and relatively high-frequency F1s (greater than 625 Hz starting frequency). Overall, error responses for a given vowel

RESPONSE

STIMULUS	beet	bit	bait	bet	Bert	bat	but	bought	bout	boat	bite	boot
beet	95											5
bit	11	64	4	4	11	1	2					3
bait	11	9	49	3	12	1	1			3	1	9
bet		8		76		8	6				2	
Bert		4	4	3	82	4			1		1	1
bat				4		67	16	2	9	2		
but	1	3		16	5	28	29	17			1	
bought				7		12	22	35	6		13	
bout		1	1			3	2		23	66	4	
boat		1	1	3	4		9	7	6	54	6	7
bite			5	6	1	1		3	4	11	69	
boot		3		7			3	2	3	1		80

FIG. 1. Percent response to synthetic "bVt" stimuli by implant patients.

were vowels with similar F1 and F2. In most instances, the error responses were not widely distributed but, rather, were the one or two vowels with the most similar formant pattern. Errors in which vowels with high F2s were heard as vowels with low F2s, for example "boot" for "beet", were relatively uncommon.

The small number of cases in which a vowel with a high F2 was confused with a vowel with a low F2 suggests that information about extreme values of F2 was available for use in identification. This information is probably specified by a place code. Electrode 2 would be maximally activated by low F2s (near 1 kHz) and Electrode 3 would be maximally activated by high F2s (near 2 kHz).

The relatively good identification of front vowels suggested that F1 was well resolved. This follows from the view that the principal cue to vowel "height" is F1. To assess this directly, four patients were tested in a same–different task to determine the difference limen for an isolated F1. The "standard" F1 was at 400 Hz, and the "comparison" F1s ranged from 400 to 500 Hz. A mean level of 75% correct was achieved with a difference of 60 Hz (Dorman et al., 1989a). This degree of resolution is sufficient for the differentiation of most F1s in English and is most probably the result of rate coding (although it is not possible from the present data to rule out an account based on the relative distribution of current along the cochlea).

Recognition of Consonants

Information about consonant identity exists in the time, amplitude, and frequency domains of the acoustic signal. If implant patients extract time-intensity (envelope) information from the signal with fidelity, then certain manner distinctions among consonants should be well marked (Summerfield, 1985). In a vowel–consonant–vowel (vCv) environment, stop consonants are marked by a period of relative silence and so should not be confused with semivowels, nasals, or fricatives, which lack the distinctive silent interval. By similar logic, the stop-initiated fricative in "chin" should not be confused with the fricatives in "sin" or "fin".

Envelope information could also contribute to the distinction between voiced and voiceless stop consonants (see, for example, Van Tassel, Soli, Kirby & Widin, 1987). In a vCv environment, voiced stops are characterized by the presence of a low-amplitude voicing signal during the period of stop closure. Voiceless stops are characterized by a period of silence during closure. However, the voiced–voiceless distinction is also marked by a number of other cues. Rosen (1989) suggests that the voiced–voiceless distinction may be signalled, most generally, by the presence of periodicity or aperiodicity, which in the Ineraid occurs most commonly in low and high channels, respectively, of the implant.

Finally, envelope cues can aid in the identification of the fricatives in "fin" and "thin" on the one hand, and in "sin" and "shin" on the other. The two pairs of voiceless fricatives are normally marked by large differences in the amplitudes of the fricative noises—the fricatives in "fin" and "thin" can be 15 dB less intense than the fricatives in "sin" and "thin", which in turn may be 5–10 dB less intense than an accompanying vowel.

As we noted above, resolution of cues in the frequency domain is constrained by the upper limit of rate coding and by limited information about place of cochlear stimulation. Most generally, these constraints will limit the resolution of cues to *place* of articulation. As a consequence, we should expect that information about place will be transmitted with less fidelity than information about manner or voicing.

The results of a consonant identification task for 10 implantees who had relatively good word identification scores are shown in Figures 2 and 3

	RESPONSE															
STIMULUS	b	d	g	p	t	k	s	sh	th	ch	z	m	n	w	l	y
b	100															
d	7	60	33													
g		20	80													
p				60	33	7										
t				7	80	7	7									
k					20	80										
s							100									
sh							7	93								
th	20	13					40		20		7					
ch										100						
z	7	7				7			13		47	7				13
m									7			53	27	7		7
n												20	60		7	13
w												20	13	47	13	7
l												7		27	60	7
y											13					87

FIG. 2. Percent response to "vCv" stimuli by "better" patients.

RESPONSE

STIMULUS	b	d	g	p	t	k	s	sh	th	ch	z	m	n	w	l	y
b	51	20	6	11	3								3	3	3	
d	14	48	28			6							3			
g	3	31	48	6	3	9										
p				46	43	6		3		3						
t				23	28	46					3					
k	3			11	26	57	3									
s	9	6	3	17	6		26		14		6	6			9	
sh								97		3						
th	20	3	3	9			17		37			3	3	3	3	
ch					6		11			83						
z	6						3		3		11			60	14	3
m	6	3							3		3	60		6	20	
n												14	57	14		14
w									3					74	17	6
l									6			3		60	23	9
y																100

FIG. 3. Percent response to "vCv" stimuli by "poorer" patients.

(Dorman et al., 1990). The stimuli were naturally produced consonants in "vCv" format.

The mean score, averaged over all subjects, was 58% correct. Three subjects achieved scores better than 60% correct. The confusion matrix for these "better" subjects is shown in Figure 2. Seven subjects achieved scores less than 60% correct. The confusion matrix for these "poorer" subjects is shown in Figure 3.

The results were analysed in term of conditional information transmitted (Soli, 1990) for envelope features of the signals and for the more traditional features of place, manner, and voicing. The envelope features were those defined by Van Tassel et al. (1987). Voicing envelope cues distinguish voiced from voiceless consonants and were described earlier in this section. Amplitude envelope cues distinguish voiced consonants with a high amplitude, /mnwrly/, from all other consonants. Burst amplitude cues distinguish voiceless stops from other consonants on the basis of the high intensity, burst/affrication segment at signal onset.

The analyses indicate that the better patients received 75% or more of the information about the features of burst envelope, amplitude envelope, voicing envelope, and frication. In contrast, the poorer subjects received only 47 to 62% of the available information about those features. Thus, the better subjects received about 30% more information than the poorer subjects for the burst envelope, amplitude envelope, and frication features. Much smaller differences (10 to 17%) were seen for the features voicing envelope, place, and nasality (see Tye-Murray & Tyler, 1989, for another feature analysis of consonant identification by Ineraid patients).

EXPERIMENT 1

Cues to Stop Consonant Place of Articulation

In the study described above, stop consonant place of articulation was recognized at an above-chance level of performance. Three cues influence the identification of place: The onset frequency and direction of the second and third formant transitions (Liberman, Delattre, Cooper, & Gerstman, 1954), the "tilt" of the spectrum at signal onset [high frequency emphasis for /d/, "peaked", middle frequency emphasis for /g/ and low frequency emphasis for /b/ (Blumstein, Issacs, & Mertus, 1982)], and the abruptness of spectral change following onset (abrupt for /b/ and /d/, non-abrupt for /g/) (Kewley-Port, 1983; Lindholm, Dorman, Taylor, & Hannley, 1988).

Our aim in Experiment 1 was to assess the contribution of these cues to the identification of stop consonants by our patients and by a group of normal-hearing listeners. To accomplish this, we created a set of stimuli in which the three stimulus factors were crossed (3 types of formant transition × three types of spectral tilt; and 3 types of formant transition × two levels of abruptness).

Method

Subjects. The subjects were 18 normal-hearing adults and three implantees. The normal-hearing listeners ranged in age from 22 to 56 years. The implantees ranged in age from 31 to 53 years. The implantees were chosen for participation in the experiment on the basis of consonant recognition scores greater than 50% correct (on the 16-item test described in the previous experiment) and on availability for testing. Patient 1 lost his hearing following a skull fracture and had been deaf for 1 year before implantation. Patient 2 lost her hearing following a progressive hearing loss of unknown origin. She had been deaf for 28 years before implantation. Patient 3 lost his hearing following Meniere's Disease. He had been deaf for 5 years before implantation.

Stimuli. The prototype stimuli were the spoken words "bill", "dill", and "gill". The stimuli with conflicting transitions and spectral shape at onset were created by digitally filtering the onsets of the signals so that the onset spectrum was similar to that of one of the two other stimuli (e.g. the stimulus with /b/ transitions was given a rising, or /d/-like, spectral shape at onset). The stimuli with conflicting transition and abruptness of frequency change cues were created by digitally concatenating or removing waveform segments at signal onset. The duration of the burst and aspiration segments of "bill" and "dill" were approximately 9 msec in duration. To create "bill" and "dill" with non-abrupt frequency change following onset, three 9-msec segments were concatenated. The burst and aspiration segment of "gill" was 27 msec in duration. To create "gill" will an abrupt frequency change following onset, an 18-msec segment was removed from the signal following the initial 9-msec segment.

The test was a randomized sequence of eight tokens of each stimulus. A familiarization sequence consisted of five repetitions of the "bill", "dill," and "gill" stimuli with a normal complement of cues. The tests were conducted in a sound-attenuated booth. Signal presentation was by loudspeaker at 75 dB SPL.

Results

The results for stimuli constructed with conflicting formant transition and spectral tilt cues are shown in Table 1. For the normal-hearing subjects an inappropriate spectral tilt did not significantly alter identification responses [the difference in percent "d" responses to the stimulus with d-formants/d-tilt and to the stimulus with d-formants/b-tilt was not significant ($t=0.44$, $p=0.74$)]. Thus, formant transition cues in this vowel context appear to determine identification responses for normal listeners. For the implantees, spectral tilt had a far greater influence. For each patient, spectral tilt altered identification accuracy. For example, for Patient 1 the stimulus with b-transitions/b-tilt was identified as "b" on 62.5% of the trials and as "d" on 37.5% of the trials. In contrast, the stimulus with b-transitions/d/tilt was identified as "d" on 62.5% of the trials and as "g" on 37.5% of the trials. The stimulus was never identified as "b". A similar outcome was reported by Dorman et al. (1988) for another Ineraid patient. It appears, then, that the spectrum of the signal at onset plays a larger role in determining consonant identification for these implant patients than for normal-hearing listeners.

The results for stimuli constructed with conflicting formant transitions and abruptness of frequency change following signal onset are shown in Table 2. For the normal-hearing subjects variation in the abruptness dimension had no effect on the identification of "b" and "g". However, significantly more "d" responses were elicited by the stimulus with non-

TABLE 1
Percent Response as a Function of Stimulus Factors[a] and Hearing Status

Stimulus Factor		Hearing Status: Normal			Patient 1			Patient 2			Patient 3		
trans.	*tilt*	*b*	*d*	*g*	*b*	*d*	*g*	*b*	*d*	*g*	*b*	*d*	*g*
b	b	98	1	0	62	37	0	37	62	0	62	25	12
b	d	96	3	1	0	62	37	0	62	37	62	25	12
b	g	99	0	1	0	75	25	0	0	100	12	87	0
d	d	25	74	1	50	37	12	12	37	50	25	75	0
d	b	34	65	1	100	0	0	50	25	25	50	25	25
d	g	26	73	1	0	37	62	0	62	37	50	25	25
g	g	2	0	98	0	12	87	0	0	100	0	0	100
g	b	6	1	92	75	12	12	37	25	37	12	62	25
g	d	0	1	99	0	0	100	0	25	75	12	37	50

[a] Formant transitions and spectral tilt.

abrupt frequency change than the stimulus with abrupt frequency change ($t = 2.75$, $p = 0.013$). For the implantees, the abruptness of frequency change cue played a larger role in identification. When the non-abrupt spectral change for /g/ was made abrupt, all patients evidenced fewer /g/ responses. In the extreme instance, S3 dropped from 100% /g/ responses to 12.5% /g/ responses. When the abrupt onset for /b/ was lengthened, two of the three patients (1 and 3) evidenced a large reduction in the number of /b/ responses and a large increase in the number of /g/ responses.

TABLE 2
Percent Response as a Function of Stimulus Factors[a] and Hearing Status

Stimulus Factor		Hearing Status: Normal			Patient 1			Patient 2			Patient 3		
trans.	*abrupt.*	*b*	*d*	*g*	*b*	*d*	*g*	*b*	*d*	*g*	*b*	*d*	*g*
b	a	98	1	0	62	37	0	37	62	0	62	25	12
b	na	97	1	2	12	50	37	37	50	12	0	25	75
d	a	25	74	1	50	37	12	12	37	50	62	37	0
d	na	13	81	5	50	0	50	0	75	25	25	75	0
g	a	1	1	96	0	50	50	12	37	50	50	37	12
g	na	2	0	98	0	12	87	0	0	100	0	0	100

[a] Formant transitions and abruptness of frequency change following onset.

Discussion

Overall, we find that spectral tilt at signal onset and the abruptness of frequency change following signal onset affect the identification responses of implantees more than the identification responses of normal-hearing listeners. It is reasonable to assume that the crude quantization of frequencies in the F2/F3 region makes formant tracking a very difficult endeavour for implantees. For this reason, the spectrum at signal onset may play a more important role for implantees than for normal-hearing listeners. In similar fashion, the temporal cue of abruptness of frequency change following onset may play a more important role for implantees than for normal-hearing subjects.

EXPERIMENT 2

The Nature of Acoustic/Phonetic Categories

In the experiments described in this section we used the technique of generating a series of stimuli that varied along a small number of acoustic dimensions, in order to assess the ability of implant patients to use details of temporal and frequency cues for the purpose of phonetic categorization. We chose two continua for study based on the consonant and vowel identification ability of patients in previous experiments. Our data on consonant identification suggested that the voicing characteristic of stop consonants was well recognized. Therefore, we constructed a series of synthetic stimuli that varied in voice-onset time (VOT) to determine whether the acoustic/phonetic boundary between voiced and voiceless stops was positioned normally. Our data on vowel identification suggested that low-frequency F1s were well resolved. Therefore, we constructed a series of stimuli in which F1 and F2 co-varied in 20-Hz steps, to determine whether the acoustic/phonetic boundary between /ɪ/ (the vowel in "bit") and /ɛ/ (the vowel in "bet") was positioned normally.

Method

Subjects. Ten normal-hearing undergraduates and six implantees served as subjects. The implantees ranged in age from 24 to 61 years. For five patients the etiology of deafness was unknown. For another the etiology was Meniere's Disease. The consonant identification scores (16 consonants in "vCv" environment) ranged from 52 to 81% correct. The vowel identification scores (12 synthetic vowels in "bVt" environment) ranged from 22 to 79% correct. The word recognition scores of the implantees ranged from 48 to 100% correct.

Stimuli. To create the VOT continuum, the syllable /ga/ was synthesized using the Kewley-Port implementation of the Klatt algorithm (Klatt, 1980). The interval between the release burst and the onset of voicing in the first formant was varied over the range −10 msec (voicing occurred before the burst) to +70 msec (voicing was delayed 70 msec). For the stimuli in which voicing onset lagged the burst, aspiration excited the upper two formants for the duration of the voice onset time. The stimuli with short intervals between burst and onset of voicing were intended to be heard as /ga/. The stimuli with long intervals were intended to be heard as /ka/.

To create a continuum in which the frequency of steady-state formants varied, the vowel /ɪ/ was first synthesized. The frequency of F1 was 420 Hz, and the frequency of F2 was 1780 Hz. F3–F5 were fixed at 2570 Hz, 3300 Hz, and 3850 Hz, respectively. Stimulus duration was 150 msecs with a 25-msec rise–fall time. Nine stimuli were created by covarying F1 and F2, in 20-Hz steps, over the range 420–580 Hz and 1780–1620 Hz, respectively. The stimuli with low F1 and high F2 were intended to be heard as /ɪ/. The stimuli with high F1 and low F2 were intended to be heard as /ε/.

Procedure. Each test was presented as a forced-choice identification test with the alternatives listed in orthographic form at the top of the printed test sheet. The test sequence for each continuum consisted of eight tokens of each stimulus presented in random order. Each test sequence was preceded by a familiarization sequence in which the extreme stimuli from the continuum were presented as stimulus pairs. Eight repetitions of the pairs were presented. All tests were conducted in a sound-attenuated booth. Signal presentation level was 75 dB SPL. The stimuli were presented through a loudspeaker.

Results

The mean identification responses of the normal-hearing listeners (± 1 *SD*) are shown as dotted functions in Figures 4 and 5. The identification functions of the implantees are shown as solid functions.

As shown in Figure 4, the mean VOT boundary (i.e. the location of 50% identification responses) for normal-hearing listeners was 42 msec, with a range of 34 to 50 msec. The VOT boundaries for Patients 1, 2, 3, and 6 were within the normal range. One patient (4) evidenced a shorter VOT boundary than normal (30 msec). Patient 5 showed consistent labelling of voiced signals, but inconsistent labelling of most of the signals that would normally have been heard as voiceless.

As shown in Figure 5, the mean boundary along the /ɪ/ − /ε/ continuum for normal-hearing listeners occurred at Stimulus 5 with a range of 4.0 to 6.5. Three patients (1, 2, 5) evidenced normal boundaries. For two patients (3, 4), the boundary was slightly beyond the normal range. The remaining patient (5) evidenced a normal boundary when the ascending portion of the labelling function was fit with a straight line, but was clearly unsure of the category affiliation of Stimuli 4–7. Patient 2 also evidence a shallow identification function. We infer that for these two patients the phonetic identity of the stimuli in the middle range of the continuum was unclear.

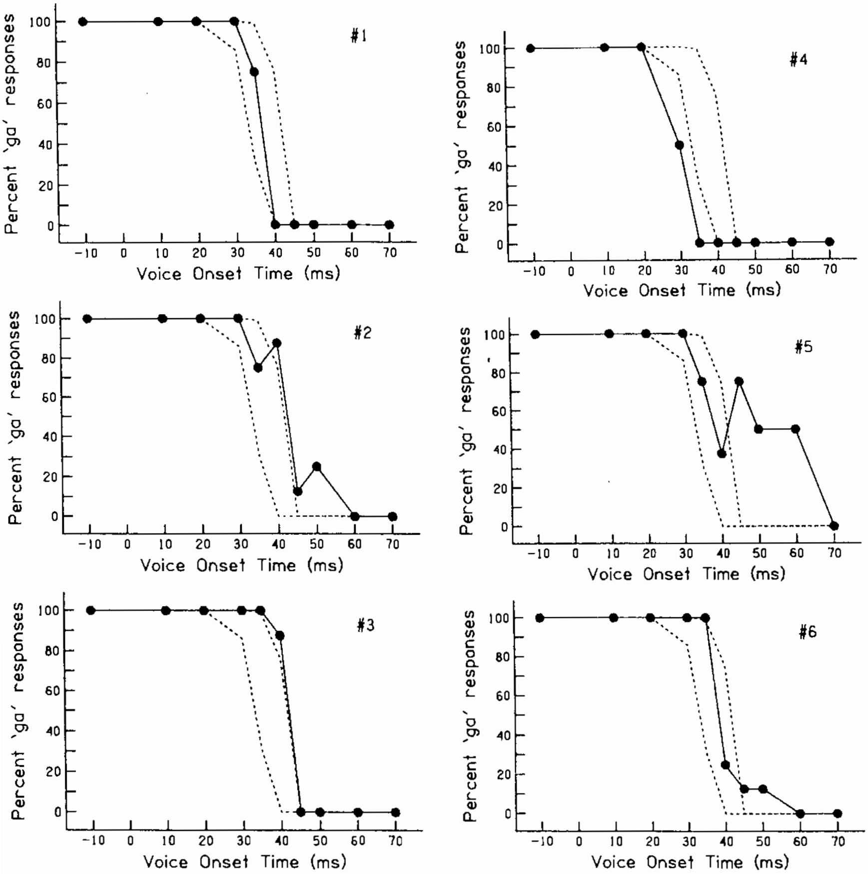

FIG. 4. Percent "ga" responses to stimuli that varied in VOT. Dotted functions indicate normal performance ± 1 *SD*.

Discussion

Two patients evidenced normal categorization, in terms of acoustic/phonetic boundaries and in slope of the identification function, of both continua. Two others evidenced boundaries only slightly outside the limits of normal. This outcome suggests that at least some implantees can appreciate the spectral/temporal detail available in the auditory representation of speech signals.

More patients were able to label the signals from the VOT continuum in a normal manner than signals from the /ɪ/-/ɛ/ continuum. This outcome probably reflects the greater number of acoustic cues available to the patients for the voicing contrast (see our discussion of cues to voicing, above).

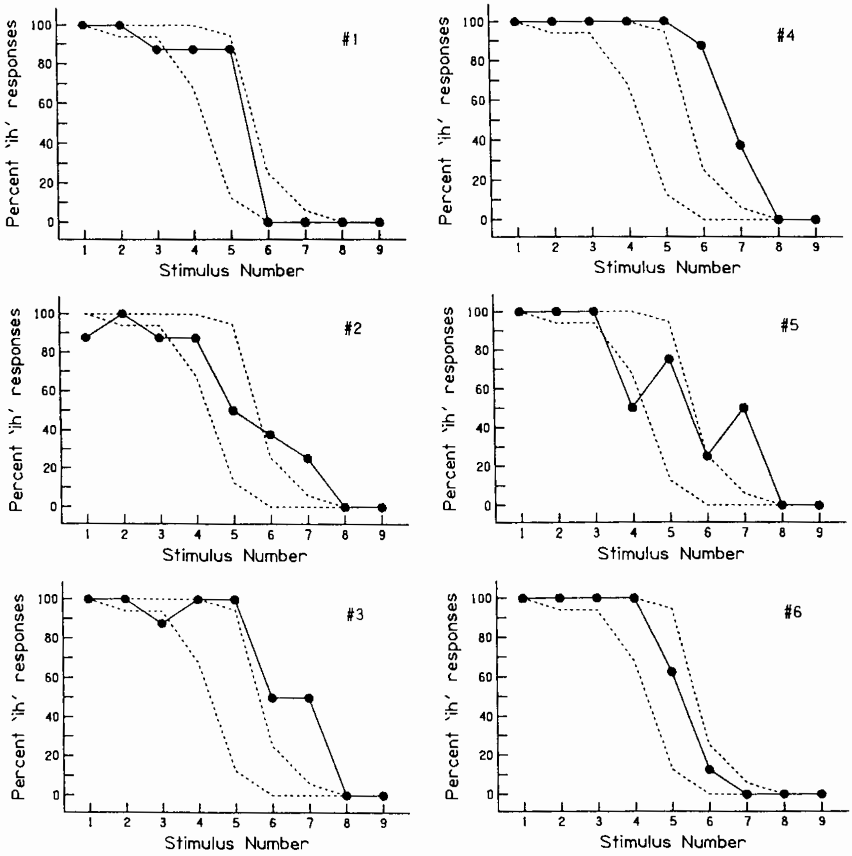

FIG. 5. Percent /ɪ/ responses to stimuli that co-varied in F1 and F2 frequency. Dotted functions indicate normal performance ± 1 *SD*.

GENERAL DISCUSSION

Our aim in this report was to describe the information that enables patients who are "good" users of a multichannel cochlear implant to identify vowels and consonants. The experiments on vowel recognition suggest (1) that F1 is relatively well resolved, and (2) that at least large differences in F2 are resolved. The relatively good resolution of F1 is due, most probably, to rate coding of F1, although, as we noted previously, the present data do not rule out the possibility of F1 coding by the relative distribution of current among electrodes. The poorer resolution of F2 is due, most probably, to the limited amount for information available about place of stimulation for frequencies above 1 kHz.

The results of our experiments on consonant recognition indicate that the best patients extract most of the amplitude envelope information (including periodicity/aperiodicity) from the acoustic signal. This information allows accurate identification of consonant manner and voicing. For consonants, as for vowels, spectral information is limited. We suspect that rate coding adds to the identification of manner and voicing by providing information about the onset frequency of F1 (Dorman et al., 1988) and, in the case of the semivowels, about the onset frequency of some F2s. Place coding provides at least some information about the onset spectra of stop consonants. The identification of stop consonant place of articulation is assisted further by the characteristic spectral/temporal cues that follow signal onset.

REFERENCES

Blumstein, S., Issacs, E., & Mertus, J. (1982). The role of gross spectral shape as a perceptual cue to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America, 72*, 43–50.

Dankowski, K. & McCandless, G. (1990). *Relationship between electrical and acoustical dynamic range and measures of speech discrimination.* Paper presented at the Second International Cochlear Implant Symposium, Iowa City, Iowa.

Dorman, M.F., Dankowski, K., Smith, L., & McCandless, G. (1989a). Identification of synthetic vowels by patients using the Symbion multichannel cochlear implant. *Ear and Hearing, 10*, 40–43.

Dorman, M.F., Dankowski, K., Smith, L., & McCandless, G. (1989b). Word recognition by 50 patients fitted with the Symbion multichannel cochlear implant. *Ear and Hearing, 10*, 44–49.

Dorman, M.F., Hannley, M., McCandless, G., & Smith, L. (1988). Acoustic/phonetic categorization with the Symbion multichannel cochlear implant. *Journal of the Acoustical Society of America, 84*, 501–510.

Dorman, M.F., 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.

Dubno, J. & Dorman, M.F. (1987). Effects of spectral flattening on vowel identification. *Journal of the Acoustical Society of America, 82*, 1503–1511.

Dynes, S. (1989). *Temporal discharge patterns in the electrically stimulated auditory nerve in cat.* Masters Thesis, Department of Physics, Massachusetts Institute of Technology.

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

Gantz, B., Tyler, R., Knutson, J., Woodworth, G., Abbas, P., McCabe, B., Hinrichs, J., Tye-Murray, N., Lansing, C., Kuk, F., & Brown, C. (1988). Evaluation of five different cochlear implant designs: Audiologic assessment and predictors of performance. *Laryngoscope, 98*, 1100–1106.

Greenwood, D. (1990). A cochlear frequency position function for several species—29 years later. *Journal of the Acoustical Society of America, 87*, 2592–2605.

Hartmann, R. & Klinke, R. (1990). Response characteristics of nerve fibers to patterned electrical stimulation. In J.M. Miller & F.A. Spelman (Eds.), *Cochlear implants: Models for the electrically stimulated ear* (pp. 135–159). New York: Springer-Verlag.

Hochmair, E. & Hochmair-Desoyer, I. (1983). Percepts elicited by different speech-coding strategies. In C.W. Parkins & S.W. Anderson (Eds.), *Cochlear prostheses* (pp. 268–279). New York: New York Academy of Sciences.

Hochmair-Desoyer, I., Hochmair, E., Burian, K., & Stiglbrunner, H. (1983). Percepts from the Vienna cochlear prosthesis. In C.W. Parkins & S.W. Anderson (Eds.), *Cochlear prostheses* (pp. 295–306). New York: New York Academy of Sciences.

Hochmair-Desoyer, I., Hochmair, E., & Stiglbrunner, H. (1985). Psychoacoustic temporal processing and speech understanding in cochlear implant subjects. In R.A. Schindler & M.M. Merzenich (Eds.), *Cochlear implants* (pp. 291–303). New York: Raven Press.

Kewley-Port, D. (1983). Time varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *73*, 322–335.

Klatt, D. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*, 971–995.

Liberman, A.M., Delattre, P., Cooper, F., & Gerstman, L. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*, *68* (8, Whole No. 379), 1–13.

Lindholm, J., Dorman, M.F., Taylor, B.E., & Hannley, M. (1988). Stimulus factors influencing the identification of voiced stop consonants by normal-hearing and hearing-impaired adults. *Journal of the Acoustical Society of America*, *83*, 1608–1614.

Rosen, S. (1989). Temporal information in speech and its relevance for cochlear implants. In B. Fraysee & N. Cochard (Eds.), *Cochlear implant: Acquisitions and controversies* (pp. 3–26). Basel: Cochlear AG.

Shannon, R. (1986). Temporal processing in cochlear implants. In M.J. Collins, T.A. Glattke, & L.A. Harker (Eds.), *Sensorineural hearing loss. The Reger conference* (pp. 349–367). Iowa City, IA: University of Iowa Press.

Smith, L. (1988). Clinical evaluation of pitch perception from a monopolar cochlear implant. *Biomedical Sciences Instrumentation*, *24*, 153–159.

Soli, S. (1990). Perceptual evaluation of a neurally based encoding strategy for cochlear implants. In J.M. Miller & F.A. Spelman (Eds.), *Cochlear implants* (pp. 378–404). New York: Springer-Verlag.

Summerfield, A.Q. (1985). Speech-processing alternatives for electrical auditory stimulation. In R.A. Schindler & M.M. Merzenich (Eds.), *Cochlear implants* (pp. 195–222). New York: Raven Press.

Townshend, B., Cotter, N., Van Campernolle, D., & White, R.L. (1987). Pitch perception by cochlear implant patients. *Journal of the Acoustical Society of America*, *87*, 106–115.

Tye-Murray, N. & Tyler, R. (1989). Auditory consonant and word recognition skills of cochlear implant users. *Ear and Hearing*, *10*, 292–298.

Tyler, R.S., Moore, B.C.J., & Kuk, F.K. (1989a). Performance of some of the better cochlear-implant patients. *Journal of Speech and Hearing Research*, *32*, 887–911.

Tyler, R.S., Tye-Murray, N., & Otto, S. (1989b). The recognition of vowels differing in a single formant by cochlear-implant patients. *Journal of the Acoustical Society of America*, *86*, 2107–2112.

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.

Revised manuscript received 23 October 1990