

## BEHAVIORAL TECHNIQUES IN AUDIOLOGY AND OTOLGY

# Identification of Synthetic Vowels by Patients Using the Symbion Multichannel Cochlear Implant\*

Michael F. Dorman, Korine Dankowski, Geary McCandless, and Luke Smith

Arizona State University, Tempe, Arizona [M.F.D.]; University of Utah, Salt Lake City, Utah [K.D., G. McC.]; and Symbion, Inc., Salt Lake City, Utah [L.S.]

### ABSTRACT

In this report we describe the vowel identification ability of eight patients who scored above 70% correct on a test of spondee word identification. The stimuli were 12 synthetic vowels in "bVt" format. The vowels differed in the frequency of F1, F2, and F3. The mean identification accuracy was 60% correct. Front vowels and diphthongs, /ɜː/, and /u/ were relatively well identified. The vowels in "but," "bought," and "bout," which were characterized by high F1s and low F2s were not well identified. The results are consistent with a model of recognition in which F1 is specified, with relatively good accuracy, by a rate code and in which extreme values of F2 are specified by a rate/place code.

In a recent report we described the word identification ability of 50 patients who use the Symbion multichannel cochlear implant (Dorman, Hannley, Dankowski, Smith, & McCandless, 1989). In that study, the best patients achieved scores of 90 to 100% correct on open set tests involving spondee words and words in sentences, and achieved scores of 50 to 60% correct on the NU 6 monosyllable test. This level of performance suggests that vowels and consonants are well identified by some patients. In this report we describe the vowel identification ability of 8 patients who achieved scores of 70% correct or greater on the test of spondee word identification.

**Implant design** The Symbion implant consists of (1) six 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 (Edgington, 1980, 1983). The most apical electrode is located about 22 mm from the round window. The electrodes are

spaced at 4 mm intervals. The most apical electrode (No. 1) is near the 1 kHz place in the cochlea, electrode No. 2 is near the 2 kHz place, electrode No. 3 is near the 4 kHz place, and electrode No. 4 is near the 8 kHz place. These four electrodes are activated in most patients. Each electrode is driven by a signal derived from the input signal after bandpass filtering. The center frequencies of the filters for channels 1 to 4 (most apical to most basal electrodes) are 0.5, 1, 2, and 4 kHz. The filters roll off at 6 dB/octave.

**Coding of Formant Frequency** Information about the formant frequencies of a vowel may be available in several forms. Consider, first, the information potentially available from neural elements discharging in synchrony with the time waveform, i.e., a "rate" code. Psychophysical studies suggest that, for some subjects, "rate pitch" increases with pulse rate to approximately 1000 pulses per sec (pps). For other subjects, rate pitch may saturate as low as 175 pps (Hochmair-Desoyer, Hochmair, Burian, & Stiglbrunner, 1983; Townshend, Cotter, Van Compernelle, & White, 1987). Difference limens for rate 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 very good vowel identification scores achieved by a few patients who use a single channel prosthesis, e.g., C. F. who scored 78% correct for a set of eight vowels (Hochmair-Desoyer, Hochmair, & Stiglbrunner, 1985), suggests that discharge rate can be used to encode formant frequencies. The level of vowel recognition obtained by C. F. suggests that the time waveform can provide information about F1 frequency and, possibly, some information about F2 frequency. It is difficult to estimate the resolution of F1 and F2 by C. F. because we do not know the contribution of pitch, signal amplitude and signal duration to recognition of the naturally produced vowels with which C. F. was tested.

Formant frequencies may also be encoded along a scale of relative frequency derived from the interaction of information from neural synchrony to the time waveform

\* This research was supported by grant R01 NS 53447-06 from the National Institutes of Health and by grants for travel from the College of Liberal Arts and Sciences and from the Vice President for Research, Arizona State University.

and information from place of cochlear stimulation. Edgington (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 signals presented to a given electrode, relative pitch increased in nonlinear fashion with stimulus frequency. For a given repetition rate, pitch increased in a nonlinear fashion as more basal electrodes were stimulated. In principle, then, if speech signals are bandpass filtered, as in the Symbion device, and the energy in the bands is directed to an array of intracochlear electrodes, the rank ordering of the frequency bands can be preserved. The resolution of such a system is limited by the number of electrodes that produce discriminably different percepts.

If relative frequency is used to encode formant frequency, then it would be useful to have an absolute anchor or reference for one end of the frequency scale. In the case of the Symbion processor, low frequency information available from the time waveform presented to the two most apical electrodes could serve as this anchor.

**Cues to Vowel Identity** Naturally produced vowels normally differ in pitch, overall sound pressure level, duration, and in the location of formants one, two, and three. Of these acoustic attributes, only the formant frequencies allow a unique categorization of vowel identity for a given speaker (Peterson & Barney, 1952; Peterson & Lehiste, 1960). Thus, pitch, SPL, and duration are "secondary" cues to vowel identity and formant frequency location is the "principal" cue. In the present experiment, we synthesized vowels in bVt format so that pitch and vowel length were constant across the set, and so that the overall SPL varied less than 1.5 dB across the set. Our aim in creating vowels in this manner was to ensure that the listeners were basing vowel identification on the principal cue to vowel identity.

## METHOD

### Subjects

The selection criteria for obtaining the Symbion device stipulates that all patients be at least 18 years of age with profound, bilateral, sensory hearing loss. They must be post-lingually deafened and cannot obtain significant speech benefit from conventional hearing aids. The patients must be in good health and must be psychologically stable.

The eight patients who participated in this research met the criteria described above. They were selected for participation in this research on the basis of their speech recognition scores (greater than 70% correct spondee recognition) and on the basis of availability for testing. The patients ranged in age from 21 to 58 years with a mean age of 37 years. The length of profound deafness ranged from 1 year to 28 years with a mean of 6.6 years. For six of the patients, deafness followed a progressive hearing loss of unknown origin. For one patient, deafness followed trauma to the temporal bone. For another patient, deafness followed administration of ototoxic drugs.

### Stimuli

The words "beet, bit, bait, bet, bat, bought, but, boot, Bert, bout, bite, boat" were synthesized on a PDP11/23 minicomputer

using the KLATT algorithm. The synthesizer was configured in parallel mode with cascade adjustment. Each stimulus was synthesized with five formants. Each stimulus was composed of a 5 msec /b/-burst, a 5 msec silent interval, 30 msec /b/-transitions, 90 msec vocalic nucleus, 50 msec /t/-transitions, 80 msec of silence, and a final /t/-burst of 50 msec. The center frequencies of the first three formants are shown in Table 1 and in Figure 1. F1, F2, and F3 were varied in frequency for some of the vowels in order to improve their naturalness and intelligibility. The F0 contour was the same for all signals. The peak SPL, measured at the level of the subject's input microphone, varied less than 1.5 dB across the 12 signals. The amplitude envelopes of the signals varied slightly.

### Procedure

The test consisted of a randomized sequence of 8 tokens of each stimulus. No feedback of correct responses was given. The familiarization sequence consisted of three repetitions of the 12 vowels. For this sequence, stimulus identity was indicated on the answer sheet. The interstimulus interval was 3.5 sec.

The test was conducted in a sound attenuating booth. The subjects sat approximately 1.5 m from a loudspeaker with their implanted ear oriented toward the loudspeaker. The signals were routed from a tape deck through an amplifier to the loudspeaker. Signal presentation level was 75 db SPL (C weighted).

## RESULTS

The averaged identification scores for the 12 stimuli are shown in Figure 2. The group mean score was 60% correct. The range of scores was 49 to 79% correct. Normal hearing listeners tested with the same stimuli achieved a mean score of 99% correct (Dorman, Hannley, McCandless, & Smith, 1988).

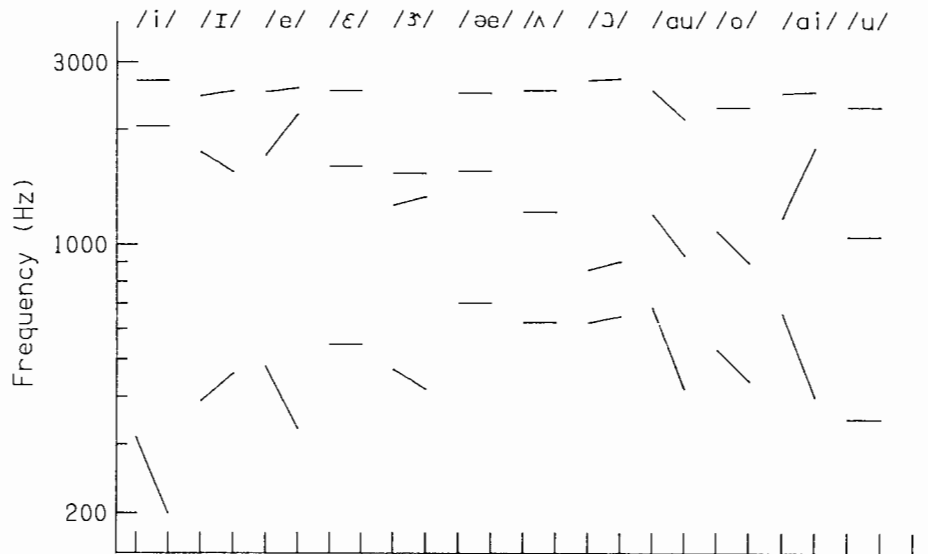
The vowels /i I ε æ/ and the diphthongs /ε ai/ were relatively well identified (mean score = 70% correct). These signals are characterized by F2s above 1500 Hz.

The vowel /3^/ was well identified (82% correct). In Figures 1 and 2 we have placed /3^/ in the front vowel series because the error matrix suggests that /3^/ was confused with front vowels. We can account for these confusions by assuming that F2 at approximately 1300 Hz and F3 at 1540 Hz combine to produce a signal in the domain of the F2s of front vowels.

**Table 1.** Center frequencies of F1, F2, and F3 for "bVt" stimuli.

Stimulus	F1	F2	F3
beet	313-200	2048	2695
bit	390-460	1756-1556	2460-2536
bet	546	1610	2539
bat	703	1562	2500
bought	625-650	859-910	2700-2730
but	625	1220	2539
boot	350	1054	2304
Bert	470-420	1270-1337	1540
boat	530-440	1088-900	2300
bout	684-420	1203-940	2538-2140
bite	660-400	1180-1880	2500-2524
bait	480-330	1720-2200	2520-2580

**Figure 1.** F1, F2, and F3 frequencies for the 12 vowels used in this experiment.



Percent response

Stimulus	Percent response												Mean of 8 patients
	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	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	

**Figure 2.** Confusion matrix for 12 synthetic vowels in bVt format.

The vowel /u/ was also well identified (80% correct). This vowel is characterized by a very low F1 (350 Hz) and a low F2 (1054 Hz).

The vowels /ʌ/ and the diphong /o/ were not well identified (mean score = 29% correct). These signals are characterized by relatively low frequency F2s (858–1200 Hz) and relatively high frequency F1s (greater than 625 Hz starting frequency).

Visual inspection of the error matrix (Fig. 2) and the formant frequencies of the vowels (Fig. 1) indicates that, most generally, the error responses for a given vowel 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

**Table 2.** Percent "different" responses to stimulus pairs with a 400 Hz "standard."

Subject	Frequency Difference (Hz)						
	0	20	40	60	80	100	120
1	0	0	80	100	100	100	100
2	0	0	0	80	80	100	100
3	10	0	10	90	100	100	100
4	0	0	0	30	90	90	100
Mean	2.5	0	22.5	75	92.5	97.5	100

frequencies. Errors in which a vowel with a low F2 was heard as a vowel with a high F2 (and visa versa) were relatively uncommon.

## DISCUSSION

The level of vowel recognition achieved by our listeners indicates that information about both F1 and F2 was available from the prostheses. We base this statement on the overall percent correct and on data from 10 normal-hearing listeners and one implant patient who were tested with bVt stimuli for which only F1 was synthesized (Dorman et al, 1988). The normal listeners averaged 39% correct for the F1-only stimuli. The patient, who achieved 79% correct for the bVt stimuli, achieved a score of 33% correct for the F1-only stimuli.

**Encoding of F2** The relatively small number of cases in which a vowel with a relatively high F2 was confused for a vowel with a relatively low F2 indicates that information about extreme values of F2 was available via the prosthesis. This information may be specified by a composite rate/place code since electrode 2 would be maximally activated by low F2s and electrode 3 would be maximally activated by high F2s. The distinctiveness of relatively low F2s may be further enhanced by rate coding of frequencies near 1 kHz.

**Encoding of F1** The relatively good identification of front vowels and the diphthongs /e ai/ suggests that F1 frequency was well resolved. This inference follows from the observation (1) that these signals are classified, in articulatory terms, as front vowels which differ in height and (2) that F1 is the principal cue to vowel height. To obtain converging evidence on our inference of good resolution of F1, we determined, for four of our patients, the difference limen for frequency of an isolated F1 resonance. The standard was a 400 Hz "F1" and the comparison stimuli were "F1s" of 420, 440, 460, 480, and 500 Hz. Each stimulus pair was presented 10 times. The order of the pairs was randomized. The results are shown in Table 2. A criterion of 75% correct was reached when the standard and comparison stimuli differed by 60 Hz. This degree of resolution is sufficient to specify the location of the first formant of front vowels.

The frequency of F1 is most likely specified by a rate code derived from the temporal waveform presented to the two most apical electrodes. It is unlikely that the resolution of F1 implied by our results could be achieved by a place code using monopolar electrodes.

**When Recognition is Poor** The poor identification of vowels with a low F2 and a high F1 may be due to several factors. Most generally, we would expect that resolution of F1 frequency, if based on a rate code, would become poorer as F1 frequency increased over the range 200 to 700 Hz. The three signals which were identified with the least accuracy had similar, relatively high frequency F1s—625 Hz for "but," from 625 to 650 Hz for "bought," and from 684 to 420 Hz for "bout." Poor resolution of similar, high frequency F1s combined with poor resolution of low frequency F2 would lead to poor recognition of "but," "bought," and "bout."

**The Information Available from Place of Stimulation** An analysis of error responses to vowels with F2s near 1 kHz bears on the issue of how information about place of cochlear stimulation is used to encode formant frequency. Recall that electrode 2 is near a 2 kHz place of stimulation but is driven by a complex waveform composed principally of frequencies around 1 kHz. If the frequency of F2 is given by the 2 kHz place of stimulation, then the vowels in "bought," "but," and "bout" should be heard as the vowels in, for example, "bet" or "bait." We base this statement on listening to signals produced by a speech synthesizer configured to produce the vowels in "bought," "but," and "bout" with an appropriate F1 frequency but a fixed F2 frequency of 2 kHz. Inspection of Figure 2 indicates that "but" was heard as "bet" on 16% of the trials, "bought" was heard as "bet" on 7% of the trials, and "bout" was heard as "bait" on 1% of the trials. The most common error responses to "but," "bought," and "bout" were words with vowel F2s in the domain of 1 kHz, i.e., "bat" for "but," "but" for "bought," and "boat" for "bout." Thus, the absolute place of cochlear stimulation does not appear to play a major role in vowel identification by our patients.

## Implant Design

Finally, our data are relevant to the issue of cochlear implant design. The data allow a comparison of vowel identification by patients fitted with two very different prostheses: The Symbion prosthesis which uses analog, simultaneous excitation of four monopolar electrodes, and the Nucleus prosthesis (Tong, Clark, Seligman, & Patrick, 1980) which uses pulsatile, nonsimultaneous excitation of 22 bipolar electrodes. Blamey, Dowell, Brown, and Clark (1987) report 64% accuracy for 7 patients fitted with the version of the Nucleus prosthesis which extracts F0, F1, and F2. The 8 patients in our study averaged 60% correct. Interpretation of the similar mean scores in the two studies is hindered by differences in the nature of the stimuli (natural versus synthetic signals) and in subject selection procedures. Tyler, Tye-Murray, and Moore (in press) have overcome these problems by testing, with the same four synthetic vowels, a sample of "better" patients fitted with the Symbion and Nucleus prostheses. The mean score for 10 patients fitted with the Nucleus device was 94% correct. The mean score for 9 patients fitted with the Symbion device was 85% correct. In both groups, two patients achieved 100% accuracy. The data from our study, from Blamey et al. and from Tyler et al indicate that similar levels of vowel recognition can be achieved via prostheses with extreme differences in speech encoding strategy and in electrode design.

## References

- Blamey PJ, Dowell RC, Brown AM, and Clark GM. Vowel and consonant recognition of cochlear implant patients using formant estimating speech processors. *J Acoust Soc Am* 1987;82:48-57.
- Dorman MF, Hannley M, McCandless G, and Smith L. Auditory/phonetic categorization with the Symbion multichannel channel cochlear prosthesis. *J Acoust Soc Am* 1988;84:501-510.
- Dorman MF, Hannley M, Dankowski K, Smith L, and McCandless G. (1989) Word recognition by 50 patients fitted with the Symbion multichannel cochlear implant. *Ear Hear* 1989;10:44-49.
- Eddington D. Speech discrimination in deaf subjects with cochlear implants. *J Acoust Soc Am* 1980;68:885-891.
- Eddington D. Speech recognition in deaf subjects with multichannel intracochlear electrodes. In Parkins CW, and Anderson SW, Eds. *Cochlear Prostheses*. New York: New York Academy of Sciences, 1983,241-258.
- Hochmair-Desoyer I, Hochmair E, Burian K, and Stiglbrenner H. Percepts from the Vienna cochlear prosthesis. In Parkins CW and Anderson SW, Eds. *Cochlear Prostheses*. New York: New York Academy of Sciences, 1983,295-306.
- Hochmair-Desoyer I, Hochmair E, and Stiglbrenner H. Psychoacoustic temporal processing and speech understanding in cochlear implant patients. In Schindler RA, and Merzenich MM, Eds. *Cochlear Implants*. New York: Raven Press, 1985,291-303.
- Peterson G and Barney H. Control methods used in a study of the vowels. *J Acoust Soc Am* 1952;2:175-184.
- Peterson G and Lehiste I. Duration of syllabic nuclei in English. *J Acoust Soc Am* 1960;32:693-703.
- Tong YC, Clark GM, Seligman PM, and Patrick JF. Speech-processing for a multiple-electrode cochlear implant hearing prosthesis. *J Acoust Soc Am* 1980;68:1897-1899.
- Townshend B, Cotter N, Van Campenolle D, and White RL. Pitch perception by cochlear implant subjects. *J Acoust Soc Am* 1987;82:106-115.
- Tyler RS, Tye-Murray N, and Moore B. (1989). Synthetic two-formant vowel perception by some of the better cochlear-implant patients. *Audiology* (in press).

**Acknowledgment:** This research would not have been possible without the gracious hospitality of C. L. Mower.

Address reprint requests to Michael F. Dorman, Ph.D., Department of Communication Disorders, Arizona State University, Tempe, AZ 85287.

Received July 30, 1988; accepted September 13, 1988.