

The recognition of vowels produced by men, women, boys, and girls by cochlear implant patients using a six-channel CIS processor

Philipos C. Loizou^{a)}

Department of Applied Science, University of Arkansas at Little Rock, Little Rock, Arkansas 72204-1099

Michael F. Dorman

Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102 and
University of Utah Health Sciences Center, Salt Lake City, Utah 84132

Verelle Powell

Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102

(Received 22 July 1996; revised 22 September 1997; accepted 21 October 1997)

Five patients who used a six-channel, continuous interleaved sampling (CIS) cochlear implant were presented vowels, in two experiments, from a large sample of men, women, boys, and girls for identification. At issue in the first experiment was whether vowels from one speaker group, i.e., men, were more identifiable than vowels from other speaker groups. At issue in the second experiment was the role of the fifth and sixth channels in the identification of vowels from the different speaker groups. It was found in experiment 1 that (i) the vowels produced by men were easier to identify than vowels produced by any of the other speaker groups, (ii) vowels from women and boys were more difficult to identify than vowels from men but less difficult than vowels from girls, and (iii) vowels from girls were more difficult to identify than vowels from all other groups. In experiment 2 removal of channels 5 and 6 from the processor impaired the identification of vowels produced by women, boys and girls but did not impair the identification of vowels produced by men. The results of experiment 1 demonstrate that scores on tests of vowels produced by men overestimate the ability of patients to recognize vowels in the broader context of multi-talker communication. The results of experiment 2 demonstrate that channels 5 and 6 become more important for vowel recognition as the second formants of the speakers increase in frequency.
© 1998 Acoustical Society of America. [S0001-4966(98)02202-4]

PACS numbers: 43.71.Es, 43.71.Ky, 43.66.Ts [WS]

INTRODUCTION

In order to perceive speech normally, a listener must sort some physically different acoustic signals into different phonetic categories, and must sort other physically different signals into the same phonetic category. The latter circumstance arises because speakers do not have identical vocal tract geometries. One consequence of different geometries is different formant frequencies for the same phonetic segment. For example, Peterson and Barney (1952) report that, across speakers (men, women and children), F_1 for the vowel /æ/ can range from 625 Hz to 1300 Hz and F_2 can range from 1600 Hz to 2600 Hz. A signal with $F_1 = 625$ Hz and $F_2 = 1600$ Hz is heard as /æ/. A signal with $F_1 = 1250$ Hz and $F_2 = 2550$ Hz is also heard as /æ/. Another consequence of the different geometries of vocal tracts is that some signals with essentially identical formant frequencies are heard as *different* phonetic segments. For example, F_1 and F_2 frequencies of approximately 625 Hz and 1700 Hz can be heard either as /æ/ or as /ɛ/ (Peterson and Barney, 1952). This problem may be partially avoided by attention to vowel length since /ɛ/ is a "short" vowel and /æ/ is a "long" vowel (House, 1961), and/or the speaker's pitch (Miller,

1989). However, the durations of short and long vowels can overlap depending on consonantal context and speaking rate. The mechanisms which allow listeners to sort signals into the appropriate vowel categories, in spite of the complexities described above, continue to be a matter of debate (see the tutorial article by Strange, 1989).

The complexities introduced by different vocal tract geometries, and therefore different formant frequencies, into vowel recognition are of importance to researchers who study speech understanding by patients fit with cochlear implants. If we present single tokens of a small number of vowels from a single speaker to our patients for identification, then the results could be seriously misleading since most of the complexities of vowel recognition have been avoided. Yet the literature on vowel recognition by patients fit with cochlear implants, including studies by the present authors, can be characterized as using a single male speaker (real or synthetic) or, at most, a single adult male and adult female speaker.

Blamey *et al.* (1987) evaluated vowel recognition performance of 28 patients using the F_0/F_1 and the $F_0/F_1/F_2$ coding strategies of the Nucleus device. The stimuli consisted of 11 vowels in /hVd/ context presented live by one male and one female Australian English speaker. Blamey *et al.* (1987) combined the scores for the two speakers. The

^{a)}Electronic mail: loizou@ualr.edu

mean percent correct score for five patients using the $F0/F1/F2$ strategy was 57% in the hearing-alone condition.

Using a set of nine vowels, Skinner *et al.* (1991) reported an overall mean score of 62% correct by patients using the $F0/F1/F2$ processor. The nine vowels were from the Iowa laser videodisc (Tyler *et al.*, 1987) and were produced by one male speaker. Skinner *et al.* (1994) later evaluated the Spectral Peak (SPEAK) coding strategy and the MPEAK strategy for the Nucleus device with patients from three English-speaking countries. For patients from the United States and Canada, a North American vowel set was used which contained 14 vowels in /hVd/ context produced by a single North American male speaker. The Australian patients were tested with an Australian vowel set, which included 11 vowels produced by one Australian male speaker. Skinner *et al.* (1994) reported a mean vowel score of 74.8% correct for the SPEAK strategy and 70.1% correct for the MPEAK strategy.

Vowel performance with the Ineraid device has been found to be similar to that of the $F0/F1/F2$ processor of the Nucleus device. Using a set of 12 synthetic vowels in /bVt/ context, Dorman *et al.* (1989) reported a mean percent correct score of 60% with scores ranging from 49% to 79% correct. Wilson *et al.* (1990) compared the performance of patients fit with the compressed analog (CA) strategy used in the Ineraid device, with the CIS strategy developed at the Research Triangle Institute. In the vowel identification test, the stimuli consisted of eight vowels in /hVd/context produced by one male and one female speaker from the Iowa laser videodisc (Tyler *et al.* 1987). Wilson *et al.* (1990) reported a similar performance for the two strategies on vowel identification. The mean vowel score was 95% correct for the CA strategy (Ineraid) and 92% correct for the CIS strategy. The Ineraid patients in the Wilson *et al.* (1990) study were tested with the CIS strategy in the laboratory, and had no experience (only a few hours) with the new strategy. It was not clear from that study whether the patients' vowel scores would improve over time with experience. This question was addressed by Dorman and Loizou (1997) who investigated the performance of Ineraid patients on vowel identification in three different conditions: (1) with the Ineraid device; (2) within hours of fitting with the CIS processor; and (3) after one month of experience with the CIS processor. The stimuli were 13 synthetic vowels in /bVt/context. The mean percent

correct scores were 35% for the Ineraid device and 41% for the CIS processor at the time of the fitting. The difference was not significant, which was consistent with the findings of Wilson *et al.* (1990) and Boex *et al.* (1994). At one month after fitting with the CIS processor, a significant increase was observed in the mean percent correct vowel recognition (58% correct). This outcome was interpreted to mean that a period of adjustment is necessary, following the fitting of the CIS processor, in which a remapping of the vowel space occurs.

As it can be seen by the preceding literature review on vowel identification by cochlear implant patients, most of the studies have used a single adult male speaker and/or a single adult female speaker. This led us to wonder how well implant patients could identify vowels when the vowels were spoken by a large number of men, women, boy, and girl informants and were produced with widely different, and at times idiosyncratic, vowel durations.

The identification of such a vowel set is of interest not only because it would indicate how well cochlear implant patients can function under conditions of real world complexity, but also because it would give us a unique window on the mechanism underlying vowel recognition. Consider that in the present study the patients used a six-electrode array and, thus, received, at most, six independent channels of stimulation. It is of interest to know how patients can identify vowels with varying formant frequencies when the vowel spectra are represented by a small number of fixed-frequency spectral components.

In the experiments which follow we describe, in experiment 1, the level of vowel recognition obtained by patients who use a six channel, Continuous Interleaved Sampling (CIS) processor when tested with vowels produced by men, women, boys, and girls. We also describe factors which may account for the errors in identification and for individual differences in performance. In experiment 2 we test the hypothesis that channels 5 and 6 of the processor are critical for the recognition of vowels produced for women, boys, and girls, but are not critical for the recognition of vowels produced by men. Finally, we comment on how vowels can be recognized when the vowel spectrum is specified by only a small number of fixed-frequency components (or channels).

TABLE I. Biographical data on implant patients. Subject S5 cannot remember ever hearing in the ear which was implanted. Sound presented to the implanted ear is lateralized to the opposite ear.

Subject	Gender	Age (years) at detection of hearing loss	Age at which hearing aid gave no benefit	Age fit with Ineraid	Age at testing	Etiology of hearing loss	Score on H.I.N.T. in quiet	Score on NU-6 words in quiet
S1	F	7	31	33	40	unknown/ hereditary	100	80
S2	M	19	19	29	41	Cogan's syndrome	100	93
S3	F	23	48	51	57	unknown	100	71
S4	M	20	46	63	68	unknown	88	46
S5	M	5	43	48	58	unknown	92	43

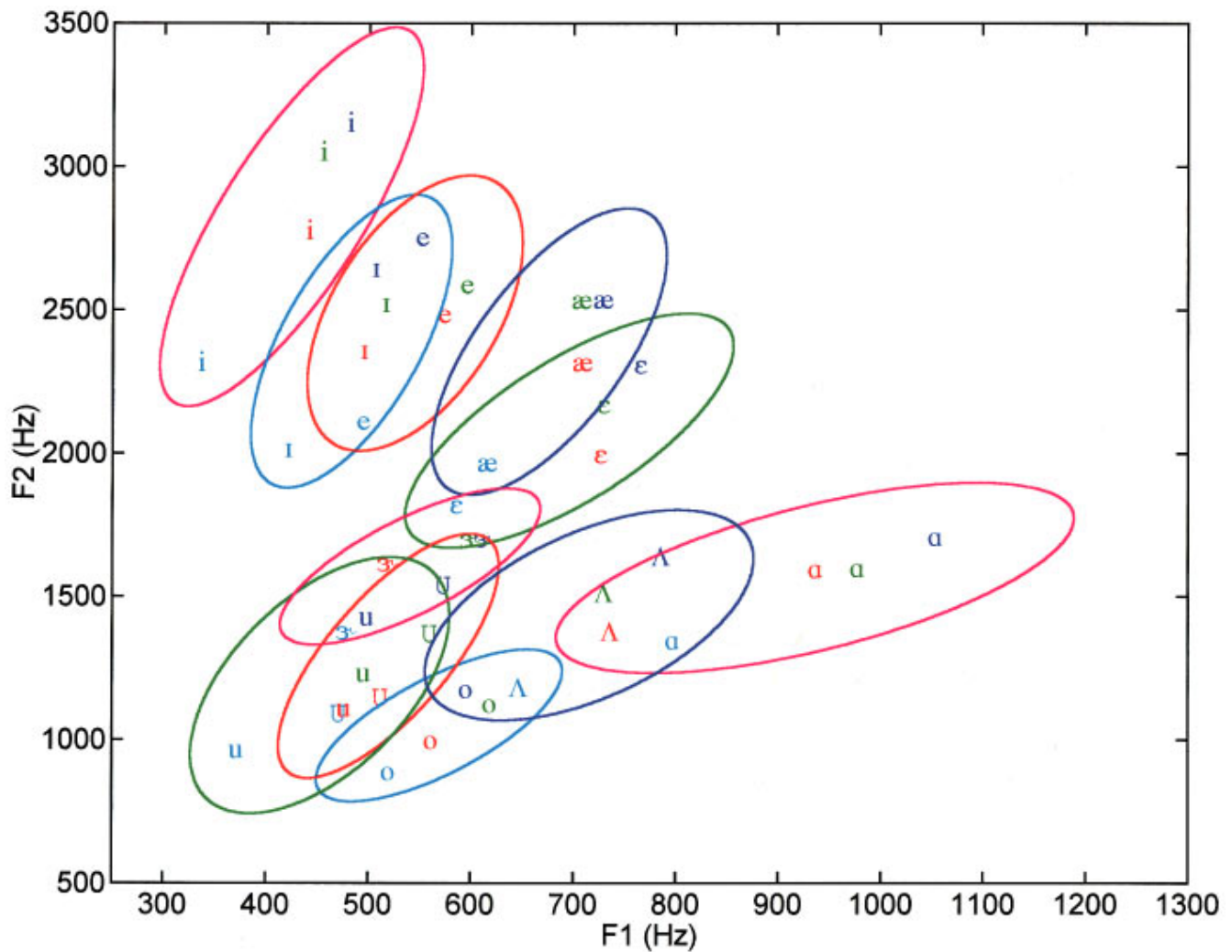


FIG. 1. Average values of $F1$ and $F2$ for men (cyan), women (red), boys (green), and girls (blue) for eleven vowels with ellipses fit to the data.

I. EXPERIMENT 1

A. Method

1. Subjects

The subjects were five post-lingually deafened adults who had used a six-channel CIS processor for periods ranging from six months to a year. All of the patients had used a four-channel, compressed-analog signal processor (Ineraid) for at least 4 years before being switched to a CIS processor. The patients ranged in age from 40 to 68 years and they were all native speakers of American English. Biographical data for each patient are presented in Table I.

2. Stimuli

The stimuli were the words “heed, hayed, hid, had, hod, head, heard, hoed, hood, hud, who’d” spoken by a total of 36 men, 36 women, 25 boys, and 19 girls. There were ten tokens of each vowel from each of the speaker categories, i.e., 10 males produced “heed,” 10 males produced “hid,” etc. In some cases the same speaker contributed tokens to several vowel categories. The stimuli were selected from recordings made by Hillenbrand *et al.* (1995). Note that the vowel /ɔ/ in “hawed” was not used in our study because it

was not identified well (only 82% correct) by the normal-hearing listeners in the Hillenbrand *et al.* study. The tokens of each vowel were selected to represent the complete area of the vowel space, i.e., the tokens came from both the center and the periphery of the vowel space. The means and distributions of the formant values of the tokens selected for this study are shown in Fig. 1. For each vowel from each speaker group the mean of the tokens used in this study were within 100 Hz of the means derived from the entire set used by Hillenbrand *et al.* (1995). This indicates that the tokens were a reasonable representation of the vowel spaces. All of the signals were identified with 90%–100% accuracy by the normal-hearing listeners in Hillenbrand *et al.* (1995).

3. Signal processing

The CIS processor was an implementation of the Wilson *et al.* (1991) design fabricated at the University of Innsbruck (Zierhofer *et al.*, 1994). The signal processor was a sixth-channel design with six-order bandpass filters (Butterworth), full-wave rectification, and a 400-Hz low-pass filter. Signals were pre-emphasized above 1200 Hz. Channel center frequencies were 393, 639, 1037, 1685, 2736, and 4443 Hz. Channel bandwidths were 187, 304, 493, 801, 1301, and

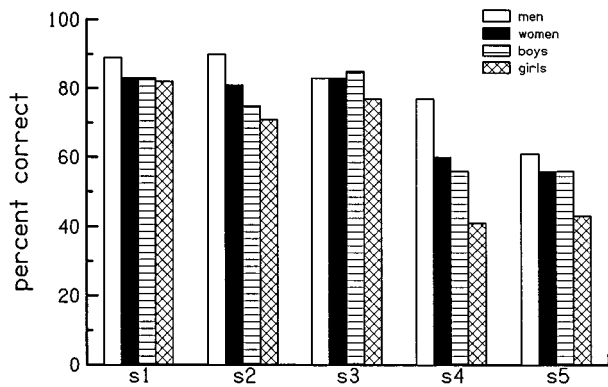


FIG. 2. Average percent correct vowel identification as a function of speaker group for five patients who used a six channel CIS processor.

2113 Hz, respectively. The channels were of equal width on a logarithmic scale. The pulse durations and pulse rates differed for each patient and were chosen after an extensive search through the pulse-duration by pulse-rate space.

4. Procedure

The test was divided into ten sessions. In each session there were three repetitions of each vowel from each of the four speaker groups. The stimuli were completely randomized within each session. Each session consisted of 132 stimuli (= 11 vowels × 4 speaker groups × 3 repetitions). A 5-min recess was allowed following each session. The test sessions were preceded by one practice session in which the identity of vowels was indicated to the listeners. In this practice block one token of each vowel from each of the speaker groups was used.

The stimuli were presented directly to the signal processors via an auxiliary input jack at a comfortable listening level. Responses were collected via a touch sensitive pad.

B. Results

The mean percent correct score for men was 80% correct; for women the mean score was 73% correct; for boys the mean score was 71% correct; and for girls the mean score was 63% correct. The data for each of the five listeners are shown in Fig. 2. A repeated measures analysis of variance indicated a significant main effect for subjects [$F(4,12) = 140.2, p < 0.0001$], a significant main effect for speaker group [$F(3,12) = 32.4, p < 0.0001$], and a significant interaction of subjects and speaker group ($F = 3.79, p < 0.0001$).

For speaker groups, *post hoc* tests according to Scheffe ($\alpha = 0.05$) indicated that (i) the men's vowels were easier to identify than vowels from the other groups, (ii) that vowels from women and boys were more difficult to identify than vowels from men but less difficult than vowels from girls, and (iii) that vowels from girls were more difficult to identify than vowels from any other group. For subjects, *post hoc* tests according to Scheffe ($\alpha = 0.05$) indicated that S1, S2, and S3 performed significantly better than S4 and S5.

The mean scores for each vowel as a function of speaker category are shown in Table II. The most difficult vowels to identify were those in "hod" (38% correct), "had" (59% correct), and "head" (62% correct). The mean score for "head" was reduced by the very poor performance on tokens from women (37% correct).

C. Discussion

In the Introduction it was noted that a test of vowel recognition using tokens from a single speaker from a single speaker group avoids the normally occurring complexities of vowel recognition. As a consequence, data collected in this fashion may overestimate the ability of cochlear implant patients to recognize vowels. The outcome of the present experiment suggests this to be the case. Vowels produced by men were easier to identify than vowels produced by women and boys, and both were easier to identify than the vowels produced by girls.

Other experiments have also found that vowels produced by men are relatively easy to identify for the best performing patients fit with six channel CIS systems. Wilson *et al.* (1990) report a mean score of 92% correct for seven "better" patients who used a CIS processor. The test set was eight vowels spoken by one man and one woman. The three "better" patients in the present experiment averaged 85% correct for the adult male and female speakers (88% and 82%, respectively). The test set included 11 vowels. Both the greater number of vowels in the set and the greater number of speakers undoubtedly contributed to the lower scores in the present experiment. At all events, the data from the present experiment suggest that vowels produced by men can be well identified with six channels of stimulation when a large, but not full, complement of vowels is tested and when the normal variation in formant frequencies for a given vowel exists within the test set.

The potential overestimation of underlying speech perception skills occasioned by the use of only male speakers is

TABLE II. Averaged identification scores (percent correct) for vowels produced by men, women, boys, and girls.

	heed	hid	hayed	head	had	hod	hoed	hood	who'd	hud	heard
Men	94	84	83	73	68	43	89	82	81	83	83
Women	84	82	81	37	55	40	93	82	81	63	92
Boys	77	85	68	75	51	37	74	77	77	63	65
Girls	78	70	67	61	60	31	62	59	65	55	62
Mean	83	80	75	62	59	38	80	75	76	66	76

TABLE III. Results of cluster analysis for distances between channel outputs of vowels produced by men, women, boys, and girls. The between-vowel variance indicates the distance between vowel categories within a speaker group. The measure of within-vowel variance indicates the variability of tokens of a given vowel within a speaker set. The Fisher ratio is the ratio of between-vowel variance to within-vowel variance.

	Between-vowel variance	Within-vowel variance	Fisher ratio
Men	7.30×10^{-2}	2.30×10^{-8}	3.56×10^6
Women	1.15×10^{-2}	1.18×10^{-8}	9.75×10^5
Boys	2.04×10^{-2}	6.20×10^{-8}	3.25×10^5
Girls	1.27×10^{-2}	4.26×10^{-8}	2.99×10^5

illustrated in Fig. 2 by a comparison of the performance of S3 and S4. For S3 performance on the vowel set produced by men (83% correct) suggests that differences in formant frequencies were relatively well resolved. This inference is confirmed by performance on the women's, boys', and girls' vowel sets (83%, 84%, and 77% correct, respectively). Consider, now, the performance of S4. He identified vowels from men with nearly the same accuracy as S3 (77% correct versus 83% correct). In this instance the inference of reasonably good resolution of formant frequencies for the other vowel sets is not appropriate since performance falls from 77% correct for men, to 60% correct for women, to 55% correct for boys, and to 41% correct for girls.

1. Accounting for the difficulty in identification of vowels from different speaker groups

As noted above, vowels produced by men were the easiest to identify. One factor contributing to this effect may be the relatively poor resolution of differences in formant frequencies for women, boys, and girls due to the increased width of the CIS processor filters into which the formants of the women, boys, and girls fall. That is, because a girl's formant frequencies will fall into higher and, therefore, wider filters than a man's, differences in formant frequencies for vowels produced by a girl should be less well resolved. To test this assumption, the signal level at the output of each processor channel for each vowel was measured.¹ To assess the distance between the channel outputs for the different vowels in the set, the results were subjected to cluster analysis (Duda and Hart, 1973) and a measure of between-class scatter was computed (see the Appendix). For this metric the larger the number, the larger the distance between vowels. Therefore, a larger distance between vowels would support the finding that the vowels are more discriminable. The results are shown in the first column in Table III. The largest value occurs for men's vowels. Women's, boys', and girls' vowels have a smaller value, indicating a smaller difference between channel outputs. These results are consistent with the outcome of better performance on men's vowels relative to women's, boys', and girls' vowels. However, the data do not predict the differences in performance on vowels produced by women, boys, and girls.²

It is also reasonable to ask whether the tokens of a given vowel are more variable within one speaker group than another. That is, it could be the case that the channel output patterns for girls, for example, are more variable for a given

TABLE IV. Dynamic range (dB) for electrodes 1–6 for patients S1–S5.

	Dynamic range (dB)					
	Ch. 1	Ch. 2	Ch. 3	Ch. 4	Ch. 5	Ch. 6
S1	15.9	19.1	19.8	22.1	19.4	16.9
S2	28.1	29.2	27.3	28.6	29.8	24.9
S3	16.1	15.7	23.5	21.8	24.5	14.5
S4	10.8	12.0	11.5	11.6	12.2	10.1
S5	15.8	16.2	18.1	18.4	10.8	7.7

vowel than for men. To assess this we used a measure of within-class scatter (see the Appendix). The results are shown in the second column in Table II where a larger number indicates greater variability. The channel output patterns of the boys and girls were found to have greater variability than those of the women and the men. This result is consistent with the poorer performance on vowels produced by boys and girls.

Finally, the measures of between-class variance and within-class variance can be combined using Fisher's ratio which computes the ratio of between-class variance to within-class variance (see the Appendix). The outcome is shown in the third column in Table III. Here a large value indicates a greater distance between items. Men's vowels had the largest value and that value was an order of magnitude larger than the value for women, boys, and girls. Girls had the smallest value. These data are consistent with the outcome that men's vowels were the easiest to identify and that the girls' vowels were the most difficult.

2. Accounting for differences among patients

As described above, subjects S1, S2, and S3 achieved higher scores than S4 and S5. A measure of auditory function—dynamic range, i.e., the range between threshold of detection for electrical stimulation and a high comfortable level of stimulation³—may be related to the better performance of patients S1, S2, and S3 relative to S4 and S5. As shown in Table IV, S1, S2, and S3, relative to S4 and S5, have, most generally, larger dynamic ranges in channels 1–4 and have much much larger dynamic ranges in channels 5 and 6. These data fit the pattern for S1, S2, and S3, relative to S4 and S5, of better recognition of male vowels (88% correct versus 70% correct) and much better recognition of vowels produced by girls (73% correct versus 41% correct). It is possible that for S4 and S5 the small dynamic ranges for electrodes 5 and 6 made coding of girls' formant frequencies more difficult.

3. Accounting for errors in the identification of a given vowel

The three most difficult vowels to identify were those in "hod," "had," and "head." Errors on "head" are the easiest to understand. As shown in Table II, the average identification score for "head" was reduced greatly by the poor identification of /ε/ produced by women. The large difference in identification accuracy among the speaker groups provided a clear window into the difficulty underlying the errors in identification of /ε/ produced by women. The aver-

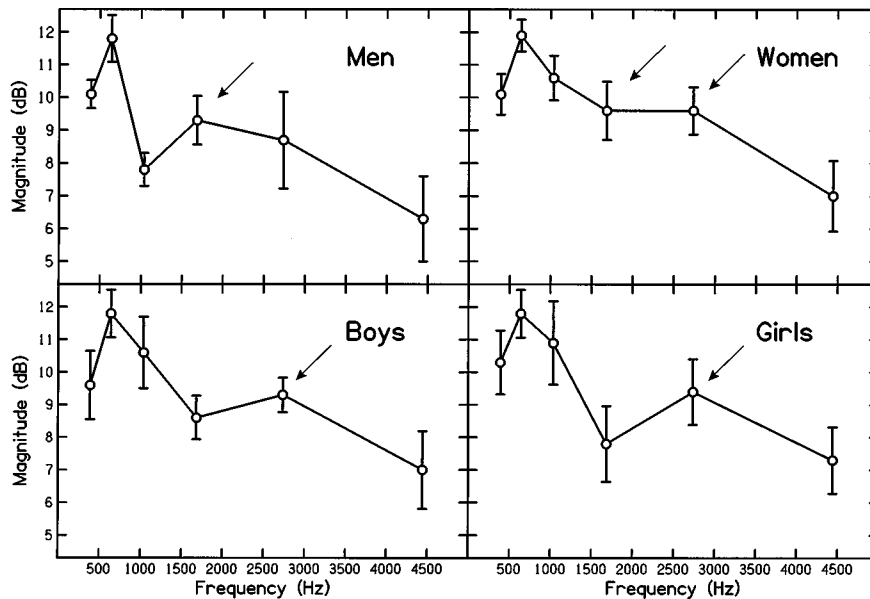


FIG. 3. Average channel output patterns for the vowel / ϵ / (in "head") as a function of speaker group. The arrows point to the channel which codes F_2 for each speaker group.

aged channel output patterns for / ϵ / produced by men, women, boys, and girls are shown in Fig. 3. The output pattern of the women speakers differs from the other patterns in the magnitude of the output for the channel coding F_2 . For men, boys, and girls there is a distinct peak in the pattern corresponding to the channel into which F_2 falls (indicated by the arrows in Fig. 3). This peak is reduced for / ϵ / produced by women. If this too low peak is responsible for the difficulty in identification, then well identified tokens of women's / ϵ / should have a larger peak than less well identified tokens of / ϵ /. As shown in Fig. 4, this was the case. Finally, inspection of the confusion matrix for vowels produced by women indicated that "head" was most often confused with "hud." We would suppose that the channel out-

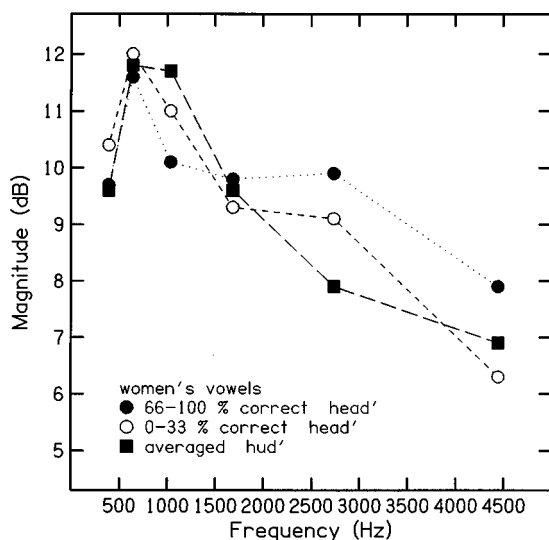


FIG. 4. Average channel output patterns for / ϵ / (in "head") and / Λ / (in "hud") produced by women.

put pattern of poorly identified tokens of / ϵ / would look much like the output pattern for / Λ /. As shown in Fig. 4, this was the case.

The vowel in "had" was the second most difficult vowel to identify. "Had" was most often confused with "head." A short account of this confusion is as follows. "Had" was produced as a diphthong; the initial portion was similar acoustically to the vowel in "head." In well identified tokens of "had" the channel output pattern during the initial portion of the syllable was similar to that in / ϵ /, while the channel output pattern during the final portion of the syllable was appropriate for / æ /. In poorly identified tokens of "had," the channel output pattern for the initial portion of the vowel was similar to that for / ϵ /, while the output pattern for the final segment was ambiguous between / ϵ / and / æ /.

The vowel in "hod" was the most difficult to identify. The two most common error responses were "had" and "hud." Curiously, there were few similarities between the channel output patterns of the poorly identified tokens of "hod" and the well identified tokens of "had" and "hud." However, the poorly identified tokens of "hod" were distinguishable from the well identified tokens, in that the poorly identified tokens lacked the distinct peak in the output pattern characteristic of the well identified tokens (Fig. 5). As shown in Fig. 5 the poorly identified tokens of "hod" were characterized by a more diffuse distribution of energy across channels 4, 5, and 6. In contrast, the well identified tokens were characterized by a significant drop in channel output level after the fourth channel. Figure 5 illustrates the distinction between the poorly identified and well identified tokens of "hod." The substitution of "had" for "hod" may also have been driven by the similar durations of the two vowels (272 ms, on average, for the vowel in "hod" and 273 ms, on average, for the vowel in "had").

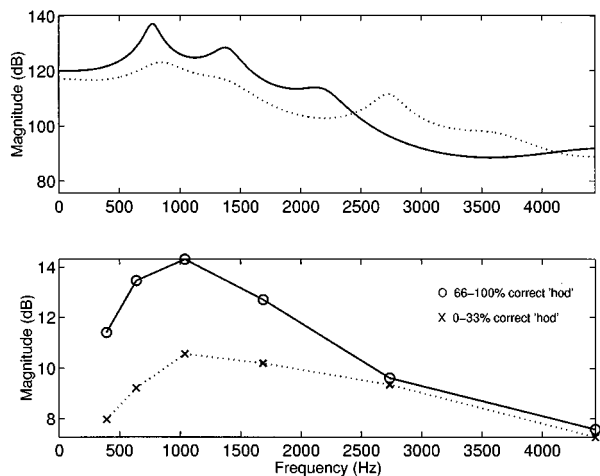


FIG. 5. (top) Spectra of the vowel /a/ (in “hod”) of a well-identified token of “hod” (solid lines) and a poorly identified token of “hod” (dashed lines) produced by male speakers. These spectra were derived using a 20-ms window, centered at the midpoint of the vowel, and a 14-pole LPC analysis. (bottom) Channel output patterns of a well-identified token of /a/ in “hod” (solid lines), and a poorly identified token of /a/ (dashed lines).

II. EXPERIMENT 2

In the discussion of experiment 1 it was suggested that channels 5 and 6 were of particular importance to the recognition of vowels produced by women, boys, and girls because of the higher formant frequencies, relative to men, produced by these speakers. To test this hypothesis, channels 5 and 6 were turned off and the patients were asked to identify vowels as in experiment 1. The hypothesis to be tested was that turning off channels 5 and 6 would have no effect on men’s vowels, but would have a significant effect on the vowels produced by women, boys, and girls.

A. Method

1. Subjects

The subjects were those described in experiment 1.

2. Stimuli

Three of the test blocks used in experiment 1 were used in this experiment. Three blocks, instead of ten, were used in order to reduce patient fatigue.

3. Signal processing

In order to eliminate channels 5 and 6, the value of the high comfortable level for channels 5 and 6 was set to the value of threshold. In this way the bandwidths of channels 1–4 were the same as those in experiment 1.

4. Procedure

The test was divided into three blocks of stimuli. In each block there were three repetitions of each vowel from each of the four speaker groups. The stimuli were completely randomized within each block. The test sessions were preceded by one practice session in which the identity of vowels was indicated to the listeners. In this practice block one token of each vowel from each of the speaker groups was used.

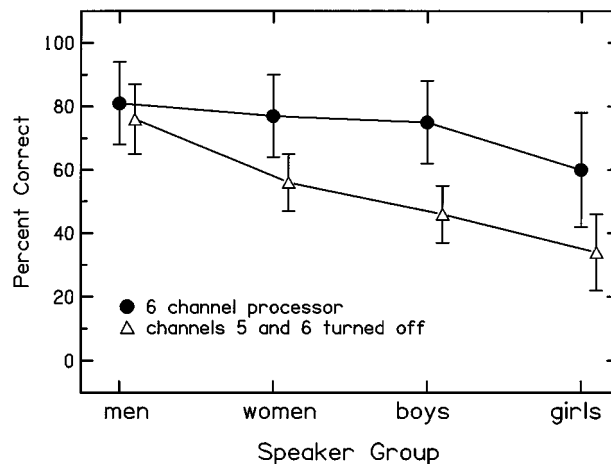


FIG. 6. Average percent correct vowel identification as a function of speaker group. The parameter is processor configuration. The solid circles indicate performance with a six-channel processor, and the open triangles indicate performance with a six-channel processor with channels 5 and 6 turned off.

As in experiment 1, the stimuli were presented directly to the signal processors via an auxiliary input jack at a comfortable listening level. Responses were collected via a touch sensitive pad.

B. Results and discussion

The mean scores with a six-channel processor and with a processor with channels 5 and 6 turned off are shown in Fig. 6. For this analysis, the three test blocks used in experiment 2 were compared to the same three test blocks used in experiment 1. A repeated measures analysis of variance indicated a main effect for speaker group [$F(3,30)=47.8$, $p<0.00001$], a main effect for channel configuration [$F(1,10)=258.4$, $p<0.00001$], and a speaker group by channel configuration interaction [$F(3,30)=11.82$, $p<0.0001$]. *Post hoc* tests according to Scheffe (alpha =0.05) indicated that the scores for vowels produced by men did not differ in the two processor conditions. The scores for women, boys, and girls did differ in the two processor conditions. This outcome, that the identification of vowels produced by men was not significantly affected by elimination of channels 5 and 6 but that the identification of vowels produced by women, boys, and girls was significantly affected, is consistent with our hypothesis that channels 5 and 6 become critical for identification as the second formants of the speakers increase in frequency.

III. GENERAL DISCUSSION

In the discussions following experiments 1 and 2, we commented (i) on differences in the recognition of vowels produced by men, women, boys, and girls, (ii) on the factors which contributed to errors in identification, (iii) on factors which might account for some patients achieving high scores while other patients achieved low scores, and (iv) on the role of channels 5 and 6 in the recognition of vowels produced by men, women, boys, and girls. In this section we suggest how

vowels with varying formant frequencies can be recognized when the spectra of the vowels are specified by only a small number of fixed-frequency components.

Consider, first, the information about vowel identity available to a normal-hearing listener. If the speaker has a low F_0 , then it is likely that harmonics in the source spectrum coincide with peaks in the vocal tract transfer function. In this instance the listener has only to detect the highest amplitude harmonics in the spectrum in order to define the formant values of the signal. However, if the speaker has a high F_0 , then there may not be harmonics which coincide with the peaks in the transfer function. In this instance, the location of the formant peak must be derived from the relative amplitudes of the harmonics which surround the peak, or from the cochlear excitation pattern which results from the pattern of harmonic amplitudes.

The cochlear implant patients in the present study were faced with a situation similar to that of normal-hearing individuals listening to signals with a very high F_0 . In this instance the “ F_0 ” was so high that only six components appeared in the spectrum. Consider, now, how six fixed-frequency components could code F_1 and F_2 for a set of vowels. Figure 3 shows the averaged, channel-output levels for the vowel in “head” spoken by men, women, boys, and girls. The differences in amplitude of the channel outputs in channels 2 and 3 code the frequency of F_1 . For men the highest output level is in channel 2 and the output of channel 3 is very low. This implies a low F_1 (mean $F_1 = 579$ Hz). For women the level of channel 3 is higher than for men indicating a higher frequency F_1 (women’s mean $F_1 = 720$ Hz). For boys the difference between channels two and three is similar to that for women, suggesting that the F_1 is the same as for women (boys’ mean $F_1 = 723$ Hz). For girls, the output of channel 3 is slightly higher than for boys indicating a slightly higher F_1 (girl’s mean $F_1 = 759$ Hz). For men, the peak in energy corresponding to F_2 is in channel 4 (mean $F_2 = 1826$ Hz). For women, channels 4 and 5 have similar levels indicating a higher F_2 than for men (women’s mean $F_2 = 2001$ Hz). For boys, the peak is at channel 5, indicating a higher F_2 than for women (boys’ mean $F_2 = 2176$ Hz), while for girls the peak is also at channel 5 but the output of channel 4 is reduced indicating a higher F_2 relative to that of boys (girls’ mean $F_2 = 2314$ Hz).

The foregoing indicates that differences in channel output levels reflect differences in formant frequencies. The results of the present experiment demonstrate that differences among the output levels of six channels can be used to code the vowel formant frequencies of male voices with reasonable adequacy. More channels will be needed to code the formants of other speakers, especially girls, with similar adequacy.

ACKNOWLEDGMENTS

This research was supported by NIDCD RO1 0000654-6. We thank Jim Hillenbrand for allowing us to use the stimuli he so laboriously gathered and analyzed. We also thank the anonymous reviewers for their helpful comments.

APPENDIX

This Appendix describes the measures used to analyze the vowel channel output vectors. These measures are based on within- and between-class scatter matrices (Duda and Hart, 1973), which are defined as follows.

Let x be a sixth-dimensional channel output vector, and m_i the mean channel vector of the i th vowel. Then, the within-class scatter matrix S_W can be defined as:

$$S_W = \sum_{i=1}^V \sum_{x \in C_i} (x - m_i)(x - m_i)^T,$$

where V is the number of vowels, and C_i is the cluster containing the channel outputs of the i th vowel. Similarly, the between-class scatter matrix S_B can be defined as follows:

$$S_B = \sum_{i=1}^V n_i (m_i - m)(m_i - m)^T,$$

where n_i is the number of channel vectors contained in cluster C_i , and m is the overall mean channel vector.

The matrices S_B and S_W can be used to measure the between- and within-vowel scatter of the channel output vectors. A simple scalar measure of the channel outputs scatter is the determinant of the scatter matrix, which provides an estimate of the hyperellipsoidal scattering volume. An estimate of the *between-vowel variance* can therefore be obtained by taking the determinant of the between-class scatter matrix S_B . Similarly, an estimate of the *within-vowel variance* can be obtained by taking the determinant of the within-class scatter matrix S_W .

The Fisher’s ratio can be constructed as the ratio of between- to within-vowel variance, i.e.,

$$F = \frac{|S_B|}{|S_W|},$$

where $|\cdot|$ denotes the determinant of a matrix. This ratio is a very popular class separability criterion in multiple discriminant analysis (Duda and Hart, 1973). Large values of F indicate that the classes, or vowels in our case, are well separated from each other, and therefore easily discriminable. Small values of F indicate that the classes are possibly intermingled with each other, and therefore easily confusable.

¹Channel output measurements were made at a point 50 ms into the time waveform. The channel outputs were computed as follows. The signal was pre-emphasized and then bandpassed into six logarithmic frequency bands using sixth-order Butterworth filters. The envelopes of the filtered signal were then extracted by full-wave rectification and low-pass filtering with a 400-Hz cutoff frequency. Six channel outputs were computed by estimating the root-mean-square (rms) energy of the six envelopes over a 10-ms frame.

²The variance metrics accounted for the differences in performance for some of the speaker groups, but failed to predict the differences in performance between other groups. This might be because the metrics used static measurements of the vowels taken at a point 50 ms into the time waveform. It is possible that if the metric had used measurements from the onset, middle, and offset segment of each vowel, that the metric might better predict the performances of all speaker groups.

³Measurements of threshold and a high comfortable level of stimulation were made with a 50-ms tone burst using a manual “up-down” procedure.

- Blamey, P., Dowell, R., Brown, A., Clark, G., and Seligman, P. (1987). "Vowel and consonant recognition of cochlear implant patients using formant-estimating speech processors," *J. Acoust. Soc. Am.* **82**, 48–57.
- Boex, C., Pellizzone, M., and Montandon, P. (1994). "Improvements in speech understanding with the CIS strategy for the Ineraid multichannel cochlear implant," in *Advances in Cochlear Implants*, edited by I. Hochmair-Desoyer and E. Hochmair (International Interscience Seminars, Vienna), pp. 136–140.
- Dorman, M., Dankowski, K., McCandless, G., and Smith, L. (1989). "Identification of synthetic vowels by patients using the Symbion multichannel cochlear implant," *Ear Hear.* **10**, 40–43.
- Dorman, M., and Loizou, P. (1997). "Mechanisms of vowel recognition for Ineraid patients fit with CIS processors," *J. Acoust. Soc. Am.* **102**, 581–587.
- Duda, R., and Hart, P. (1973). *Pattern Classification and Scene Analysis* (Wiley-Interscience, New York).
- Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**, 3099–3111.
- House, A. (1961). "On vowel duration in English," *J. Acoust. Soc. Am.* **33**, 1174–1178.
- Miller, J. (1989). "Auditory perceptual interpretation of the vowel," *J. Acoust. Soc. Am.* **85**, 2114–2134.
- Peterson, G., and Barney, H. (1952). "Control methods used in a study of vowels," *J. Acoust. Soc. Am.* **24**, 175–184.
- Skinner, M., Clark, G., Whitford, L., Seligman, P., Staller, S., Shipp, D., Shallop, J., Everingham, C., Menapace, C., Arndt, P., Antogenelli, T., Brimacombe, J., Pijl, S., Daniels, P., George, C., McDermont, H., and Beiter, A. (1994). "Evaluation of a new spectral peak coding strategy for the Nucleus 22-channel cochlear implant system," *Am. J. Otolaryngol.* **15** (Suppl. 2), 15–27.
- Skinner, M., Holden, L., Dowell, R., Seligman, P., Brimacombe, J., and Beiter, A. (1991). "Performance of postlinguistically deaf adults with the wearable speech processor (MSP) of the Nucleus multi-channel cochlear implant," *Ear Hear.* **12**, 3–22.
- Strange, W. (1989). "Evolving theories of vowel perception," *J. Acoust. Soc. Am.* **85**, 2081–2087.
- Tyler, R., Preece, J., and Lowder, M. (1987). "The Iowa audiovisual speech perception laser videodisc," Laser Videodisc and Laboratory Report (University of Iowa, Department of Otolaryngology Head and Neck Surgery).
- Wilson, B., Finley, C., Lawson, D., Wolford, R., Eddington, D., and Rabinowitz, W. (1991). "Better speech recognition with cochlear implants," *Nature (London)* **352**, 236–238.
- Wilson, B. S., Lawson, D., and Finley, C. (1990). "Speech processors for auditory prostheses," Fourth Quarterly Progress Report on NIH Project No. N01-DC-9-2401, 1–9.
- Zierhofer, C., Peter, O., Bril, S., Pohl, P., Hochmair-Desoyer, I., and Hochmair, E. (1994). "A multichannel cochlear implant system for high-rate pulsatile stimulation strategies," in *Advances in Cochlear Implants*, edited by I. Hochmair-Desoyer and E. Hochmair (International Interscience Seminars, Vienna), pp. 204–207.