

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**  
Progress Report No. 23 (1999)  
*Indiana University*

**“Vowel Spaces” of Normal-Hearing and Hearing-Impaired  
Listeners with Cochlear Implants<sup>1</sup>**

**James D. Harnsberger, Mario A. Svirsky,<sup>2</sup> Adam R. Kaiser,<sup>2</sup>  
Richard Wright,<sup>3</sup> and David B. Pisoni**

*Speech Research Laboratory  
Department of Psychology  
Indiana University  
Bloomington, Indiana 47405*

---

<sup>1</sup> This work was supported by NIH-NIDCD Training Grant DC00012, NIH-NIDCD Grant R01-DC00111, and NIH-NIDCD Grant R01-DC03937. We would like to thank Chris Quillet for his technical assistance.

<sup>2</sup> Also DeVault Otologic Research Laboratory, Department of Otolaryngology-Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, Indiana; Department of Biomedical Engineering, and Department of Electrical Engineering, Purdue University.

<sup>3</sup> Now at the University of Washington, Seattle, WA.

## **“Vowel Spaces” of Normal-Hearing and Hearing-Impaired Listeners with Cochlear Implants**

**Abstract.** Cochlear implant users show substantial individual differences in their ability to understand speech in general, and vowels in particular. One possible reason for these differences lies in their widely different abilities to identify formant frequencies. Another possible reason is that cochlear implants present spectral information to cochlear locations that are more basal than normal. The latter explanation has been controversial. Some authors have proposed that the spectral mismatch introduced by cochlear implants may be completely overcome by cochlear implant users (Rosen et al., 1999), while others believe that spectral mismatch may result in important limitations to speech perception, no matter how much time is used by the cochlear implant users to adapt to the new percepts. In the present study, we designed a vowel perception test, using a Method-of-Adjustment (MOA) paradigm, to compare the vowel spaces of eight cochlear implant users to those obtained from 43 normal hearing listeners with the same dialect as the cochlear implant users. The MOA vowel test consisted of a set of 330 steady-state synthetic stimuli arranged in a two-dimensional grid, generated by varying the first and second formants of the vowels. Subjects were asked to label and rate on a seven-point scale those stimuli which matched ten visually-presented vowel stimuli corresponding to /i/, /ɪ/, /e/, /ɛ/, /æ/, /ʌ/, /ɑ/, /u/, /ʊ/, and /o/. Two-dimensional plots of subjects' responses for all ten target stimuli constituted the “vowel spaces” of the subjects. With one exception, no systematic shift was observed across all ten vowel categories in the vowel spaces of cochlear-implant users, suggesting that these subjects were able to adapt completely to the spectral shift introduced by the implant. However, the cochlear-implant users' spaces differed substantially from normal vowel spaces in terms of the relative size of the vowel categories, and their location in perceptual space.

### **Introduction**

Although cochlear implants allow profoundly deaf people to hear, cochlear implant users show a very wide range of speech perception skills. The most successful cochlear implant users can easily hold a face-to-face conversation, and they can even communicate on the telephone, a difficult task because there are no visual cues available and because the signal itself is highly degraded. On the other hand, the least successful cochlear implant users have a difficult time communicating even in a face-to-face situation, and can barely perform above chance on auditory-alone speech perception tasks. It is important to remember that electrical hearing as provided by a cochlear implant is quite different from normal acoustic hearing. One important difference lies in the ability to discriminate formant frequencies. For example, Kewley-Port and Watson (1994) report difference limens between 12 and 17 Hz in the F1 frequency region for normal hearing listeners. In the F2 frequency region, they found a frequency resolution of approximately 1.5 percent. In cochlear implant users, discrimination of formant frequencies is dependent on two factors: the frequency-to-electrode map that is programmed in their speech processor, and the individual's ability to discriminate stimulation pulses delivered to different channels. It is not uncommon for some cochlear implant users to have formant frequency difference limens that are one order of magnitude higher than those of listeners with normal hearing, or even more. It is reasonable to hypothesize that cochlear implant users with such limited frequency discrimination skills will find it quite difficult to identify vowels accurately because formant frequencies are important cues to vowel identity. Another important difference between acoustic and electric hearing is related to the fact that cochlear implants do not stimulate the entire

neural population of the cochlea but only the most basal 25 mm at best. Therefore, cochlear implants stimulate cochlear locations that are more-basal-than-normal and thus elicit higher frequency percepts than normal acoustic stimuli. For example, when the input speech signal has a spectral peak at 300 Hz, the neurons stimulated in response to this signal may have characteristic frequencies of 1000 Hz or even higher. This represents a rather extreme modification of the peripheral frequency map, resulting in more-basal-than-normal stimulation. To the extent that CI users have enough plasticity in the auditory nervous system to successfully “re-map” the place frequency code in the cochlea, the more-basal-than-normal stimulation provided by a CI should not hinder speech perception. On the other hand, an inability to re-map the place frequency code may severely limit speech perception in CI users and may be an important source of individual differences in speech perception.

Several studies have addressed the issue of adaptation to changes in frequency-to-electrode assignments for cochlear implant users. Skinner et al. (1995) showed that users of the SPEAK stimulation strategy identified vowels better with a frequency-to-electrode table that mapped a more restricted acoustic range into the subject’s electrodes than the default frequency-to-electrode table. The experimental table that resulted in better vowel perception represented a more extreme shift in spectral information than the default table, suggesting that listeners with cochlear implants can indeed adapt to such shifts, at least within certain limits. Another study that demonstrates the adaptation ability of human listeners in response to spectral shifts was conducted by Rosen et al. (1999), who used acoustic simulations of the information received by a cochlear implant user who has a more basal spectral shift of 6.5 mm in the basilar membrane (equivalent to 1.3-2.9 octaves, depending on frequency). Initially, the spectral shift reduced word identification (1% correct, as compared to 64% for the unshifted condition), but after only 3 hours of training, subjects’ performance improved to 30% correct. This result raises the question of what may have been the maximum performance of a listener who had a better chance to reach asymptotic levels. Recently, Shannon et al. (1999) performed an experiment in which the frequency-to-electrode tables of three cochlear implant users were shifted one octave with respect to the table they had been using daily for at least three years. It is important to note that this one-octave shift was in addition to the original shift imposed by the cochlear implant. After three months of experience with the new table, it was apparent that adaptation was not complete because, on average, subjects did not reach the same levels of speech perception that they had achieved before the table change. Taken together, these studies show that auditory adaptation to a modified frequency map is possible but it may be limited, depending on the size of the spectral shift.

In the present study, we investigated the issue of spectral shift with a new paradigm, a method of adjustment (MOA) procedure. This procedure was used to map the perceptual vowel spaces of adult, postlingually deafened cochlear implant users. Similar tasks have been used with normal hearing listeners (Johnson et al., 1993). In this task, subjects select the region of the F1-F2 plane that sounds (to them) like a given vowel, and the procedure is repeated for ten English vowels. This task gives us the opportunity to simultaneously assess a cochlear implant user’s plasticity, by comparing the locations of his/her selected regions to those selected by normal hearing listeners, and his/her frequency discrimination skills, by examining the spread of his/her selected regions. More specifically, a listener who was unable to adapt to the basalward spectral shift introduced by their cochlear implant would select regions that are systematically shifted to lower frequencies with respect to the regions of the vowel space selected by normal hearing listeners. In addition to the MOA task, the ability of cochlear implant users to identify synthetic vowels that differed only in F1 was measured, as well as their ability to identify natural vowels. One long-term goal of our research is to understand the mechanisms that underlie speech perception by cochlear implant (CI) users and, in so doing, gain an understanding of the individual differences in psychophysical characteristics which may explain individual differences in speech perception with a CI. With the combination of tasks employed in the present study we hope to be able to tease out the effect of

two kinds of limitations on vowel perception by cochlear implant users: frequency discrimination and auditory plasticity.

## Methods

### Participants

Forty-three normal-hearing Indiana University undergraduates and eight cochlear-implant (CI) users, all monolingual speakers of English, participated in this experiment. The normal-hearing participants consisted of 20 males and 23 females ranging in age between 18 and 28, none of who reported any history of a speech or hearing problem. The normal-hearing participants were recruited to represent the dialect of American English spoken in central Indiana with a common inventory of vowels. Only normal-hearing listeners who reported living their entire lives in central Indiana were allowed to participate in this experiment. Central Indiana was defined in terms of a 60 mile radius around Indianapolis, roughly covering the Midland dialect as described by Wolfram and Schilling-Estes (1999), and avoiding two other regional dialects found at the northern and southern extremes of the state. These latter two regional dialects are reported to differ from the Midland dialect in terms of vowel quality and degree and type of diphthongization (Labov, 1991). For participating in two 1 hour sessions, the participants received either \$7.50/hour or two credits towards their research requirement if they were enrolled in an undergraduate psychology class.

The CI user participants were recruited from the population of patients served at the Department of Otolaryngology-Head and Neck Surgery at the Indiana University School of Medicine in Indianapolis. The demographics of the CI users are given in Table 1. All of the CI users had received implants at least one year prior to participating in this study. Five were users of the Nucleus-22 device with the SPEAK strategy, while three were users of the Clarion device with the CIS strategy. The SPEAK strategy (Skinner et al., 1994) filters the incoming speech signal into up to 20 frequency bands, which are associated with different intracochlear stimulation channels. Typically, six channels are *sequentially* stimulated in a cycle that is repeated 250 times per second. The channels to be stimulated each cycle are chosen based on the filters with the highest output amplitude. In contrast, the CIS strategy (Wilson et al., 1991) as implemented in the Clarion device filters the signal into eight bands, one for each stimulation channel. All channels are sequentially stimulated with pulses whose amplitudes are determined by the filters' outputs. The stimulation cycle is repeated at a fast rate of at least 833 times per second. The main differences with the SPEAK strategy are that CIS uses a higher stimulation rate, fewer stimulation channels, and that CIS stimulates all channels in a cycle rather than choosing the channels with more energy within the corresponding filter passbands. The CI users' age ranged from 37 to 67, averaging 59.

Subject	Age (Years)	Age at Onset of Deafness	Age at Implantation	Implant Use (Years)	Gender	Implant Type	Insertion Depth
CI1	67	43	61	6	F	Clarion 1.0	-
CI2	35	29	31	3	M	Clarion 1.2	-
CI3	37	34	36	1	M	Nucleus 24	Full (5)
CI4	74	27	71	2	F	Nucleus 22	Full (7)
CI5	63	56	57	5	M	Nucleus 22	Full (4)
CI6	70	*	66	3	M	Nucleus 22	Full (7)
CI7	68	*	65	2	F	Nucleus 22	Full (8)

CI8	58	*	52	6	M	Clarion 1.0	-
-----	----	---	----	---	---	-------------	---

**Table 1:** Demographics of subjects with cochlear implants. Insertion depth and the number of stiffening rings not inserted are noted in parentheses. \*CI4, CI8 Progressive; CI5, Progressive childhood.

## Stimulus Materials

**Method-of-Adjustment Task.** The stimulus set consisted of 330 synthetic, steady-state vowels in isolation, generated by a Klatt synthesizer, that varied from one another in their first and second formants, in equal increments of 0.377 Bark (Klatt & Klatt, 1990). The Bark increment size was chosen as a close approximation of the just-noticeable-difference for vowel formants of Flanagan (1957). The F1 and F2 values for this stimulus set ranged between 2.63 Z (250 Hz) - 7.91 Z (900 Hz) and 7.25 Z (800 Hz) - 15.17 Z (2800 Hz). These ranges were chosen to represent the full range of possible values for speakers of American English, and were successfully used in an earlier method-of-adjustment study of vowel perception (Johnson et al., 1993), as well as in piloting. All of the other synthesis parameters for this stimulus set also followed Johnson et al. (1993). The values, or the formulas for calculating the values, of the most relevant synthetic parameters are summarized in Table 2. The f0 parameter was varied to generate two sets of the 330 stimuli, one representing a male voice and one representing a female voice. All of the synthetic sounds were calibrated to a 70 dB listening level.

Parameter	Value or Formula
Duration	250 ms
f0	Male: 120 Hz over the first half, dropping to 105 Hz at the end Female: 186 Hz over first half, dropping to 163 Hz at the end
F3	$\{(0.522 * F1) + (1.197 * F2) + 57\}$ or $\{(0.7866 * F1) - (0.365 * F2) + 2341\}$ <sup>4</sup>
F4	3500 Hz or (F3+300 Hz), whichever higher
B1	$29.27 + (0.061 * F1) - (0.027 * F2) + (0.02 * F3)$
B2	$-120.22 - (0.116 * F1) + (0.107 * F3)$
B3	$-432.1 + (0.053 * F1) + (0.142 * F2) + (0.151 * F3)$
B4	200 Hz

**Table 2:** The values or formulas used for an important subset of the parameters used in generating the synthetic stimulus sets.

**Vowel Identification Task.** The vowel identification task is a closed-set task that uses 9 vowels in an “h-vowel-d” format. The stimuli were digitized from the female vowel tokens of the Iowa laserdisc (Tyler et al., 1987). Only the steady state vowels (i.e., not the diphthongs) were used. There were three separate productions of each vowel. Listeners were administered three lists that consisted of five repetitions of each vowel and three practice tokens.

**F1 Identification Task.** The stimuli for this experiment were seven synthetic three-formant vowels, with an F2 value of 1500 Hz and an F3 of 2500 Hz. F1 varied linearly, from 250 Hz for stimulus 1 to 850 Hz for stimulus 7. Three-formant vowels rather than single-formant vowels were used in order to measure

<sup>4</sup> The first formula was used for the half of the grid with higher F2 values, while the second was used for the half of the grid with lower F2 values.

place-pitch discrimination with more realistic and speech-like stimuli. The stimuli were created using the Klatt 88 (Klatt & Klatt, 1990) speech synthesizer software. Voicing amplitude increased linearly in dB from zero to steady state over the first 10 ms, and back to zero over the last 10 ms of the stimulus. Steady-state amplitude was loudness balanced within 1 dB for the seven stimuli. Total stimulus duration was 1 second. The stimuli were digitally stored using a sampling rate of 11025 Hz at 16 bits of resolution and were presented from an Intel® based PC equipped with a SoundBlaster compatible sound card. Stimuli were presented at a level of at least 70 dB C weighted SPL over an Acoustic Research loudspeaker. Custom task specific software was used to present stimuli and record responses.

## Procedures

**Method-of-Adjustment Task.** The procedures varied slightly for each participant group in the study. Normal-hearing participants were tested in a quiet room in two 1-hour sessions, with the second session taking place approximately one week after the first session. In the first session, normal-hearing participants completed the method of adjustment task with one of the synthetic stimulus sets, either the male or the female voice set. In the second session, participants completed the method of adjustment task with the remaining stimulus set. The experiment was balanced for the order in which the stimulus sets were presented to the normal-hearing participants. In contrast, the CI users were tested in a quiet room or a sound-attenuated chamber, in a single test session, varying in length by individual CI user between 1 to 3 hours. Given the length of time CI users required to complete the MOA task and other demands on their time, they were presented with only one of the two stimulus sets, the male-voice set.

Each participant was presented with a two-dimensional (15 rows and 22 columns) visual grid centered in a computer screen. The grid consisted of the 330 synthetic stimuli described above. A single word appeared above this grid, constituting the target stimulus for a given trial. The visual target stimulus for a given trial was one of ten words, “heed,” “hid,” “aid,” “head,” “had,” “who'd,” “hood,” “owed,” “odd,” and “hut,” each of which contained one of the 10 vowels under study, /i/, /ɪ/, /e/, /ɛ/, /æ/, /ʌ/, /ɑ/, /u/, /ʊ/, and /o/. Subjects were instructed to search the grid, playing out individual sounds until they found the region of the grid that played out synthetic sounds that matched the vowel in the visual target stimulus. After selecting one or more synthetic sounds that matched the target, subjects were asked to give each synthetic sound a rating on a 1 - 7 scale, grading how close a match the synthetic sound was to the target. The order of presentation of each target stimulus varied randomly from participant to participant, with a single repetition of a stimulus set presented to listeners.

The particular stimuli chosen and their respective ratings were used to calculate category “centers” and sizes for each vowel type for each listener group. Category centers were determined by averaging, in the F1 and F2 dimensions, across all selected stimuli, with their contribution to the average weighted by their rating. Category sizes in each formant dimension were based on the standard deviation of the formant mean. Normal-hearing listeners’ category centers were expected to appear in F1 X F2 space in a similar arrangement to that observed in vowel production studies with American English (i.e., Peterson & Barney, 1952). The category centers of CI patients were expected to deviate from those of normal-hearing listeners depending on the extent of their more basal stimulation, which would result in an overall space that is shifted lower in F1 and F2.

**Vowel Identification Task.** The vowel identification task was a closed-set speech perception task in which three separate productions of each of ten /hVd/ tokens were presented in random order, one at a time, and the CI users had to say which one of the ten stimuli they thought they heard by responding

verbally.<sup>5</sup> All subjects heard a total of at least 15 presentations of each vowel (except CI5 who heard 10), and were instructed to guess if they did not know which vowel was presented. The subjects' responses were tabulated and scored for total percentage of correct responses.

**F1 Identification Task.** The F1 identification task was an absolute identification task, with the stimuli labeled "1" through "7" in order of increasing F1. In this task, all stimuli were played in sequence several times, so the subjects could become familiar with the stimuli. Then, the stimuli were presented ten times each in random order and subjects were asked to identify the stimulus that was presented. The subject's response and the correct response were displayed on the computer monitor before moving on to the presentation of the next stimulus. After each block of 70 presentations (ten presentations of each of seven stimuli), the mean and standard deviations of each of the seven stimuli were calculated. The  $d'$  for each pair of successive stimuli was calculated as the difference of the two means divided by the average of the two standard deviations. These  $d'$  measurements were then cumulated to calculate a cumulative  $d'$  curve, which provided an overall measure of the subject's ability to discriminate and pitch rank the seven stimuli. To calculate the cumulative  $d'$  curves, we followed the assumption that the maximum possible value of  $d'$  was 3. The average JND (defined as the mean stimulus difference resulting in  $d' = 1$ ) was calculated based on the cumulative  $d'$  curve. Given that the F1 range spanned by the seven stimuli was 600 Hz, the JND was defined as  $(600/\text{cumulative } d')$ . At least eight blocks of 70 presentations were carried out, as these were sufficient for all subjects to reach a plateau in performance, as measured by the cumulative  $d'$ . The cumulative  $d'$  reported here is the average of the best two blocks for each subject.

## Results

### Normal-Hearing Participants

The normal-hearing listener group was expected to select vowel category centers with first and second formant values that corresponded to those typical in vowel production in American English. Figure 1 shows the mean vowel categories for all of the normal-hearing subjects, for the male-voice stimulus set. Each category center is shown along with error bars denoting two standard deviations from the category centers in both formant dimensions. In this figure, all of the ratings have been used to calculate the center and size of all ten vowel categories. Figure 2 shows the vowel spaces of normal-hearing subjects calculated using only ratings of four and above. The rating of four was chosen because it was the highest rating that still allowed for category sizes to be calculated for all ten vowels of all of the normal-hearing and CI participants. The center of each category was determined by averaging the Bark values of all synthetic stimuli that were chosen in each dimension, with the means weighted by the ratings given to individual stimuli.

The perceptual spaces shown in Figures 1 and 2 demonstrate that the method-of-adjustment technique for measuring vowel categories was successfully used to generate vowel spaces that bear the typical intervowel relationships that have been observed in F1 X F2 spaces generated from vowel production data. For instance, front vowel category centers have a higher F2 than back vowel centers; high vowel category centers have a lower F1 than low vowel category centers. While no vowel production studies have been published for English speakers from central Indiana,<sup>6</sup> Hillenbrand, Getty, Clark, and Wheeler (1995) examined the vowel production spaces of 45-48 men, women, and children who were

<sup>5</sup> If a subject's answer was unclear to the experimenter, subjects were asked to respond by pointing to the answer on a test sheet.

<sup>6</sup> However, work is in progress in our laboratory on a vowel production study of Central Indiana English, using as talkers the same normal-hearing listeners that participated in this study.

native speakers of American English, specifically the variety spoken in southern Michigan. Hillenbrand et al. (1995) was a replication and extension of the classic work on the acoustics of American English vowels carried out by Peterson and Barney (1952). Figures 3 and 4 are plots of the vowel production centers from these two studies, respectively, both including the MOA vowel category centers (calculated using ratings of 4 and above). While there are differences between the absolute locations of the MOA vowel centers and the vowel production centers of the two studies, all three vowel spaces show a common set of F1 and F2 distances between vowels. The first and second formants from the normal-hearing listeners were significantly correlated with their counterparts in both the Hillenbrand et al. (1995) set ( $r = .98$ ,  $p \leq 0.01$  for F1;  $r = .89$ ,  $p \leq 0.01$  for F2) and the Peterson and Barney (1995) set ( $r = 1.0$ ,  $p \leq 0.01$  for F1;  $r = .83$ ,  $p \leq 0.05$  for F2).

### Cochlear Implant Patients

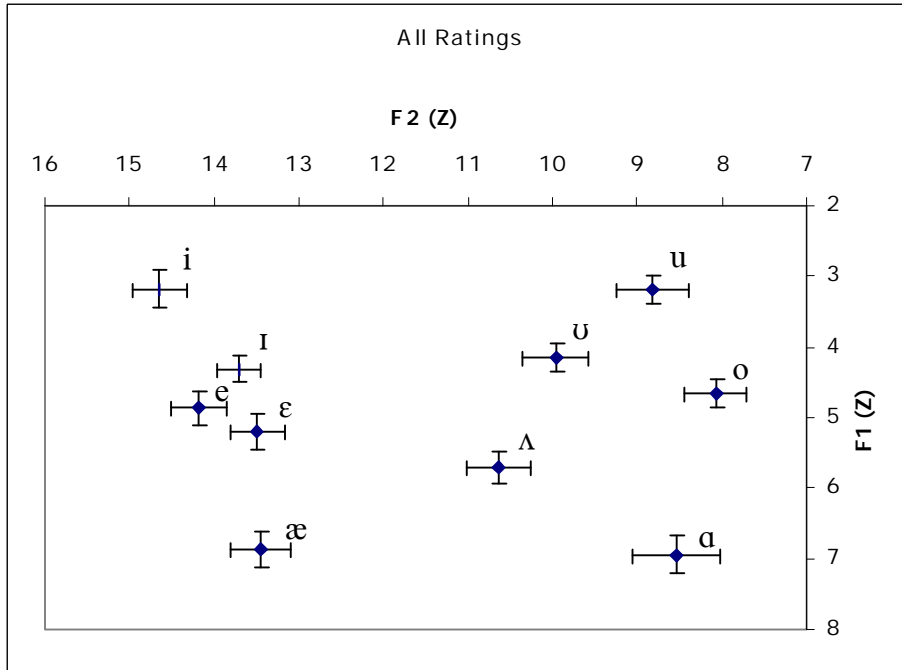
The results of the MOA, F1 identification, and vowel perception tasks for the eight CI users are shown in Figures 5 - 12. Each figure shows the labeled centers of the ten vowel categories, with the size of each category in each dimension indicated via error bars. The centers and sizes of the categories were computed via a mean in each dimension which was weighted by the rating given to each synthetic stimulus, using ratings of only four or above, just as in the normal-hearing listeners' vowel space in Figure 2. To the right of each CI user's vowel space, the results of his/her F1 identification ( $JND_{F1}$ ) and vowel perception tests (VOWEL) are listed.

An examination of all eight vowel spaces reveals little evidence of any systematic shift due to a lack of adaptation to more-basal-than-normal stimulation, with the possible exception of CI1. Instead, we find individual CI user spaces that differ from the normal space in terms of the sizes of perceptual categories, their degree of overlap, and the region of perceptual space that particular categories occupy. These observations are supported by measures of the differences between CI and normal vowel categories. Table 3 lists the absolute differences, in Bark, between the category centers of both normal and individual CI users, in both F1 and F2. Positive differences indicate that the formant of the normal-hearing listeners for that vowel was greater than the formant of the CI patient. If a shift were observable with a particular CI user, one would expect to see positive differences in one or both of the formants of the vowels of that user. Of the eight CI users, seven showed positive and negative differences, depending on the vowel and formant in question. Only one subject, CI1, systematically lowered formants for his/her category centers, as would be predicted for a listener who has not adapted completely to the spectral shift introduced by the cochlear implant (see Figure 5). We can see that CI1's categories were shifted towards the lowest formant values in the upper right corner, resulting in a particularly compressed space. However, the magnitude of this shift was not the same for all ten vowels, or for both formants. The magnitude of the shift varied from 0.28 Z-3.53 Z, with on average a greater shift observed in F1 than F2.

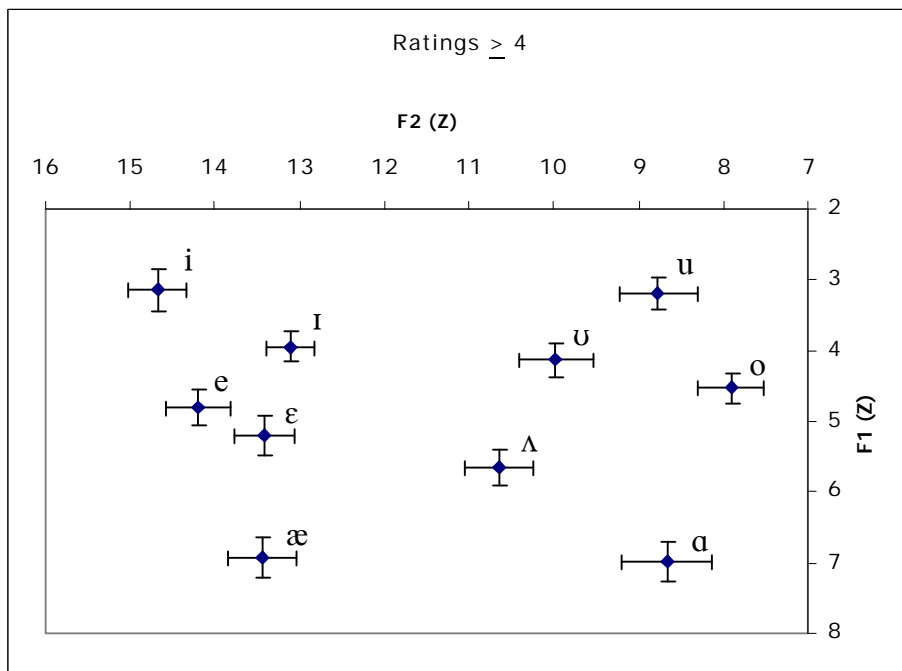
Name	F	Vowel Category										Mean
		i	ɪ	e	ɛ	æ	ʌ	ɑ	u	ʊ	o	
CI1	F1	0.43	1.06	1.79	2.37	3.53	0.62	2.82	0.45	1.32	1.52	1.59
	F2	0.28	0.8	1.09	1.46	1.46	0.39	0.57	1.41	2.35	0.66	1.05
CI2	F1	0.05	-1.7	0.02	0.14	-0.86	-1.16	-0.93	-0.52	0.07	-0.92	-0.58
	F2	-0.13	0.31	-0.77	-1.13	0.52	-0.19	0.57	0.24	1.97	-4.1	-0.27
CI3	F1	-0.62	-0.57	-1.47	-1.65	-0.75	0.62	-0.15	0.2	0.09	1.15	-0.32
	F2	0.27	0.2	0.06	0.29	1.16	0.64	0.43	-1.29	1.54	0.21	0.35

CI4	F1	-0.47	0.39	-0.35	0.08	0.09	-0.44	1.04	-0.11	0.89	0.63	0.18
	F2	0.24	-1.3	-0.59	-0.81	-0.53	-0.57	-1.83	0.46	-0.6	0	-0.55
CI5	F1	-2.01	-2.91	-0.01	-0.89	-0.31	0.29	-0.1	-0.55	-0.69	0.12	-0.71
	F2	0.05	2.08	-0.1	0.13	4.31	-0.38	-0.26	-1.07	-0.14	-2.07	0.26
CI6	F1	-2.03	-1.23	-1.16	-1.2	0.44	0.51	0.6	-0.14	-0.65	0.86	-0.4
	F2	0.3	-0.45	0.38	-1.04	0.9	0.75	-1.08	0.1	0.6	-0.38	0.01
CI7	F1	-2.09	-1.74	-0.94	-0.27	1.97	0.67	1.49	-1.25	-0.62	-0.49	-0.33
	F2	2.57	0.15	1.33	1.35	2.73	-0.27	-2.09	-2.49	-1.01	-2.15	0.01
CI8	F1	-2.18	-0.6	-0.06	-0.05	0.65	-0.08	0.93	-2.22	-1.28	-1	-0.59
	F2	3.42	1.23	2.72	1.1	1.85	-1.89	-2.47	-2.71	-2.34	-2.93	-0.2

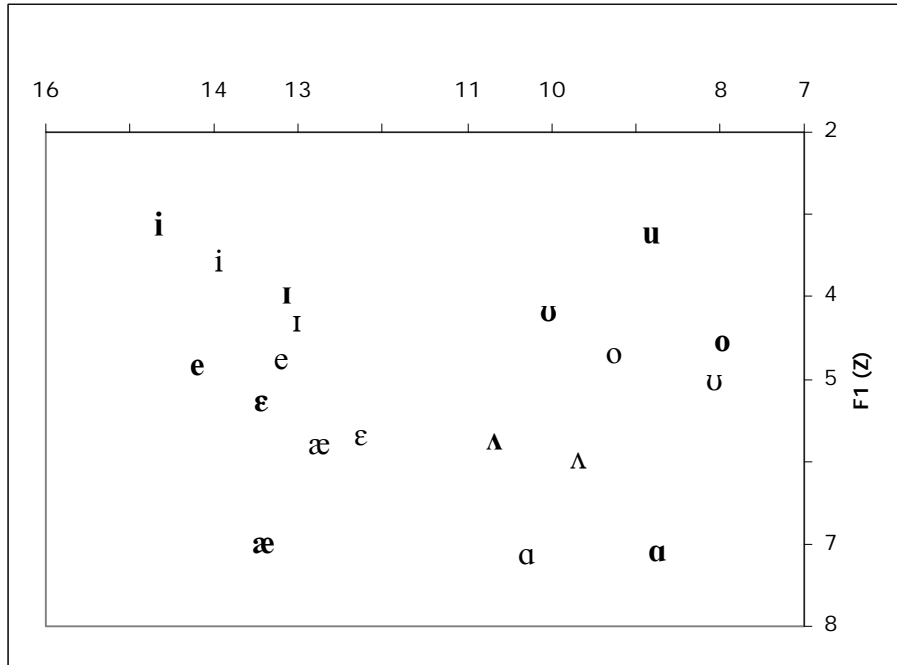
**Table 3:** The differences in Bark between normal and CI vowel categories, for each formant (F), for individual vowel categories, and the mean difference across all categories.



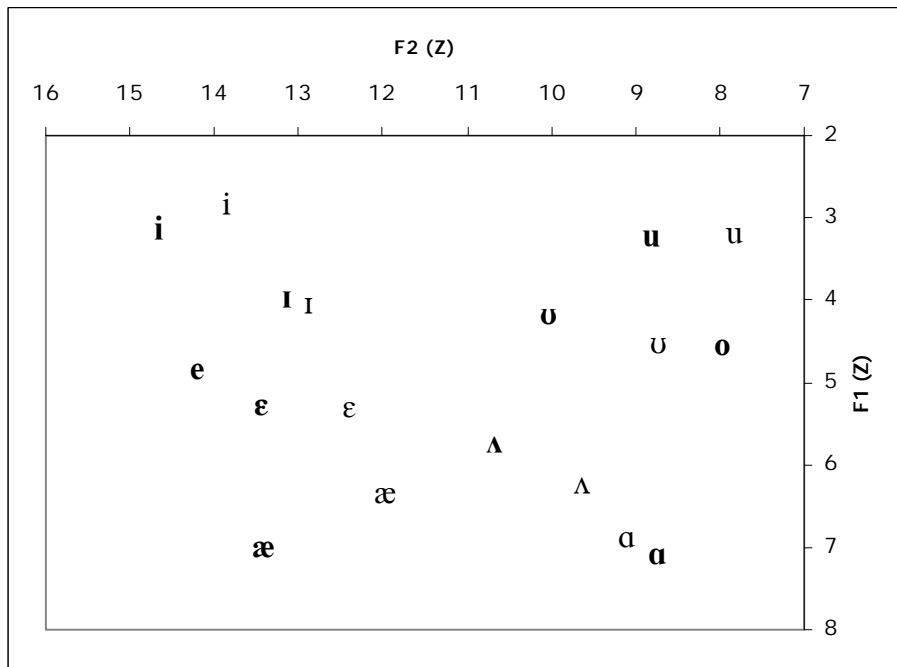
**Figure 1:** The mean vowel space of normal-hearing listeners, calculated using all of the ratings provided.



**Figure 2:** The mean vowel space of normal-hearing listeners, calculated using only ratings of four or above.



**Figure 3:** A comparison of the vowel category centers from the MOA task (in bold) with the vowel production centers (in plain text) from Hillenbrand et al. (1995).



**Figure 4:** A comparison of the vowel category centers from the MOA task (bold) with the vowel production centers (plain text) from Peterson & Barney (1952).

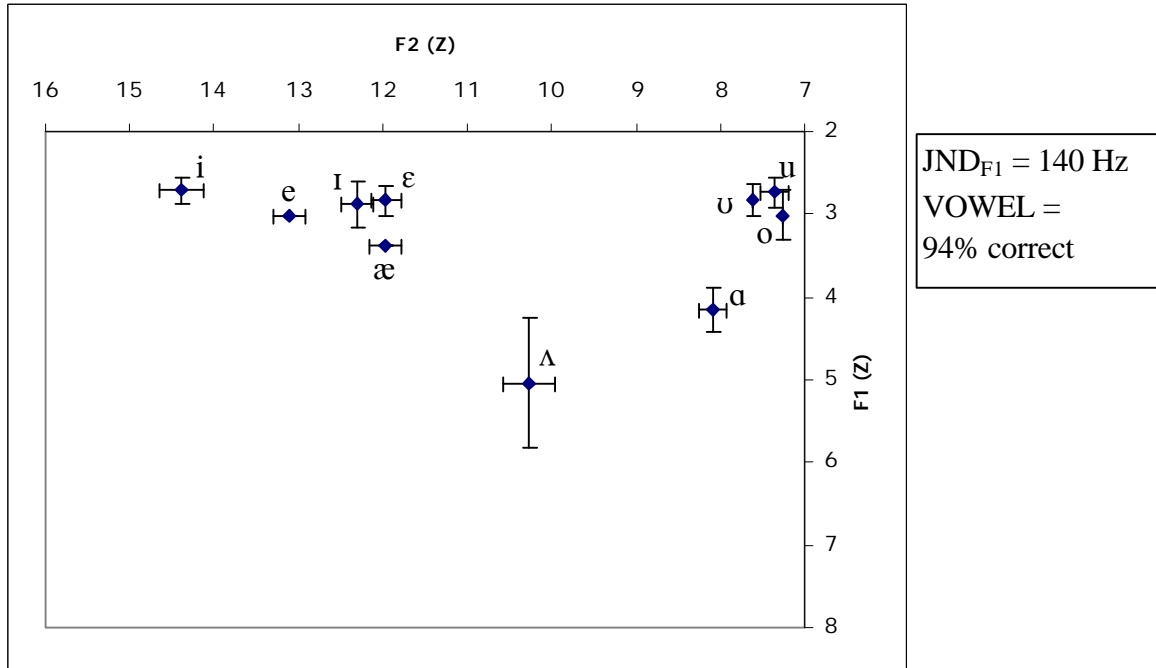


Figure 5: The vowel space of subject CI1.

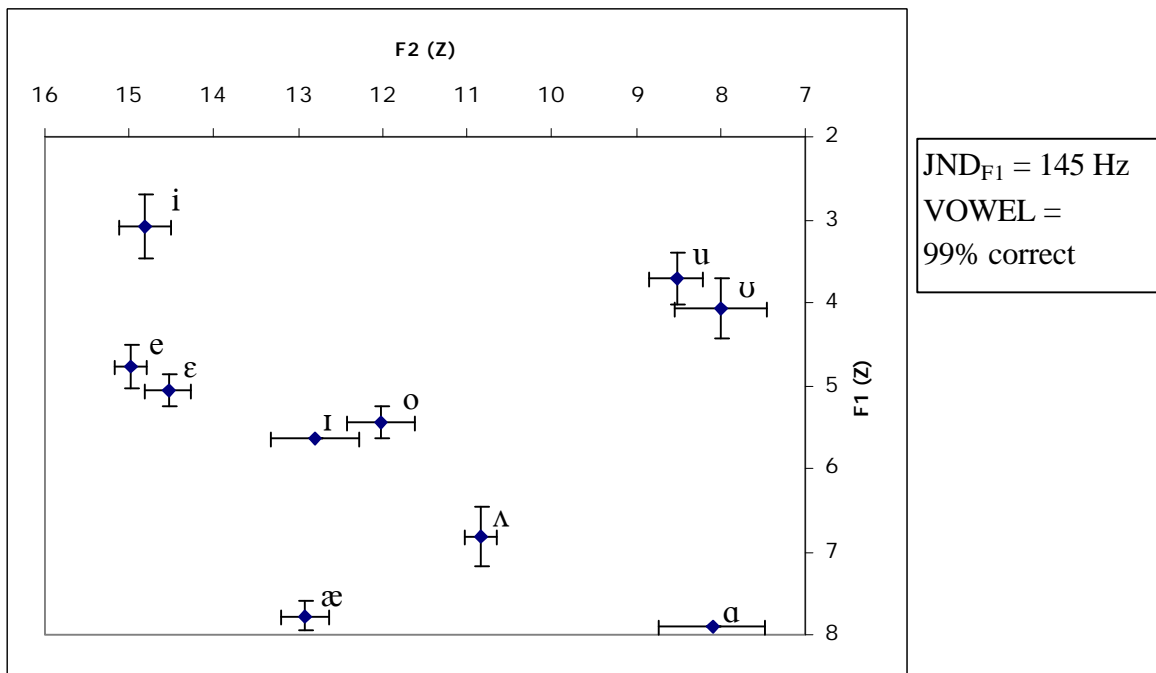


Figure 6: The vowel space of subject CI2.

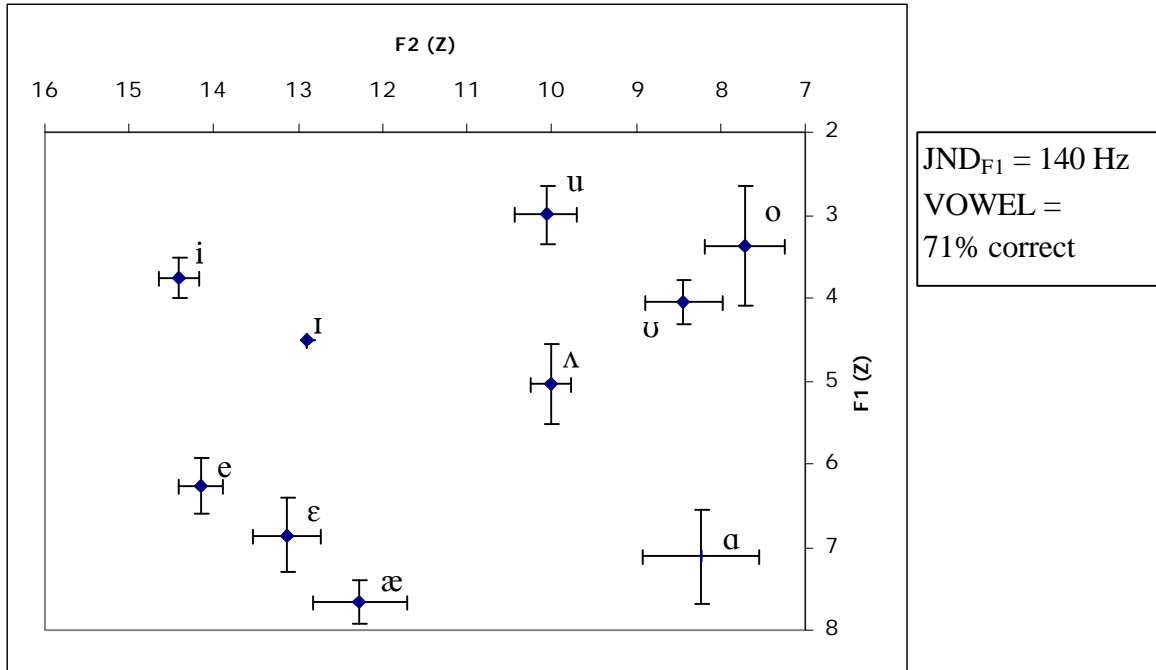


Figure 7: The vowel space of subject CI3.

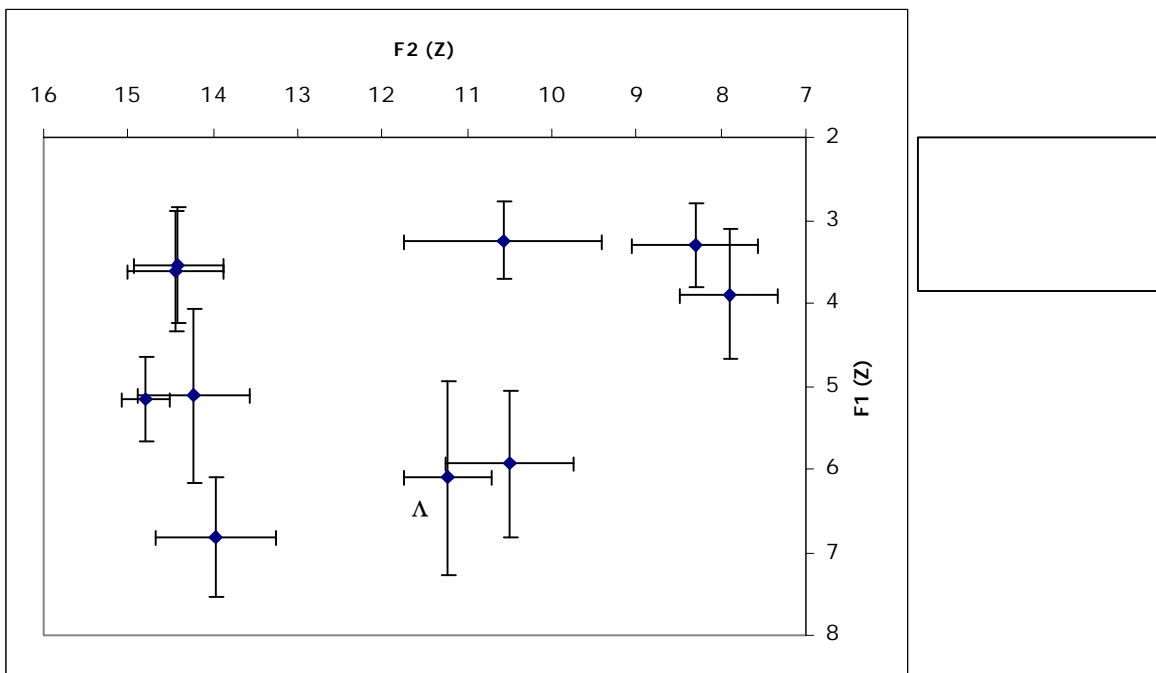


Figure 8: The vowel space of subject CI4.

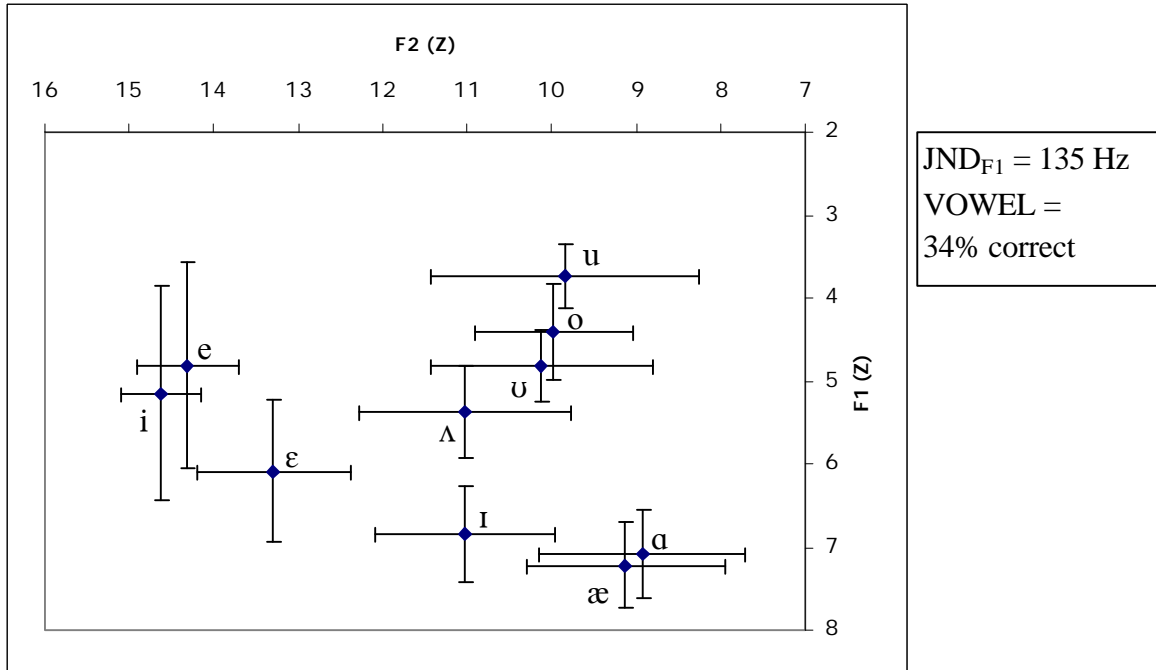


Figure 9: The vowel space of subject CI5.

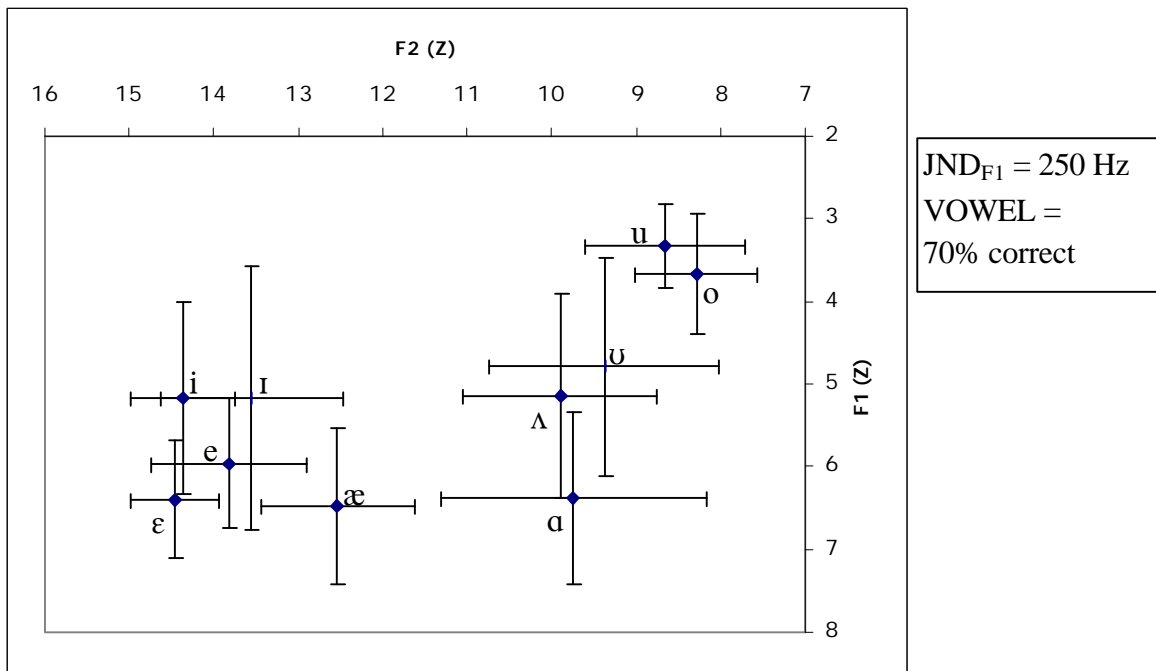


Figure 10: The vowel space of subject CI6.

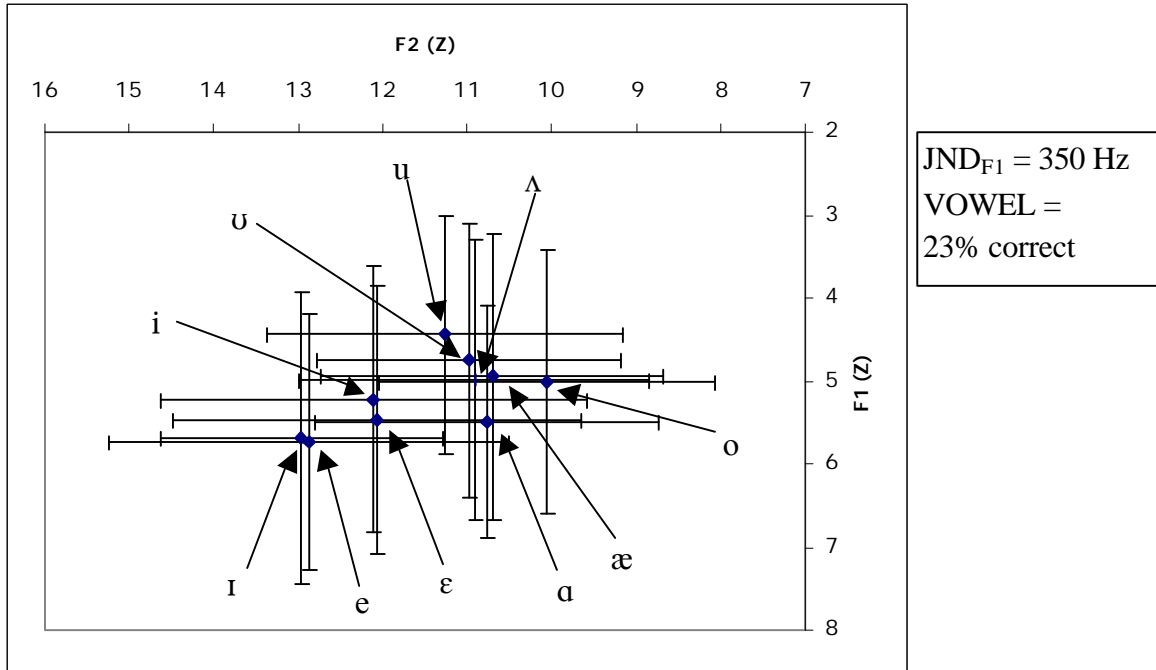


Figure 11: The vowel space of subject CI7.

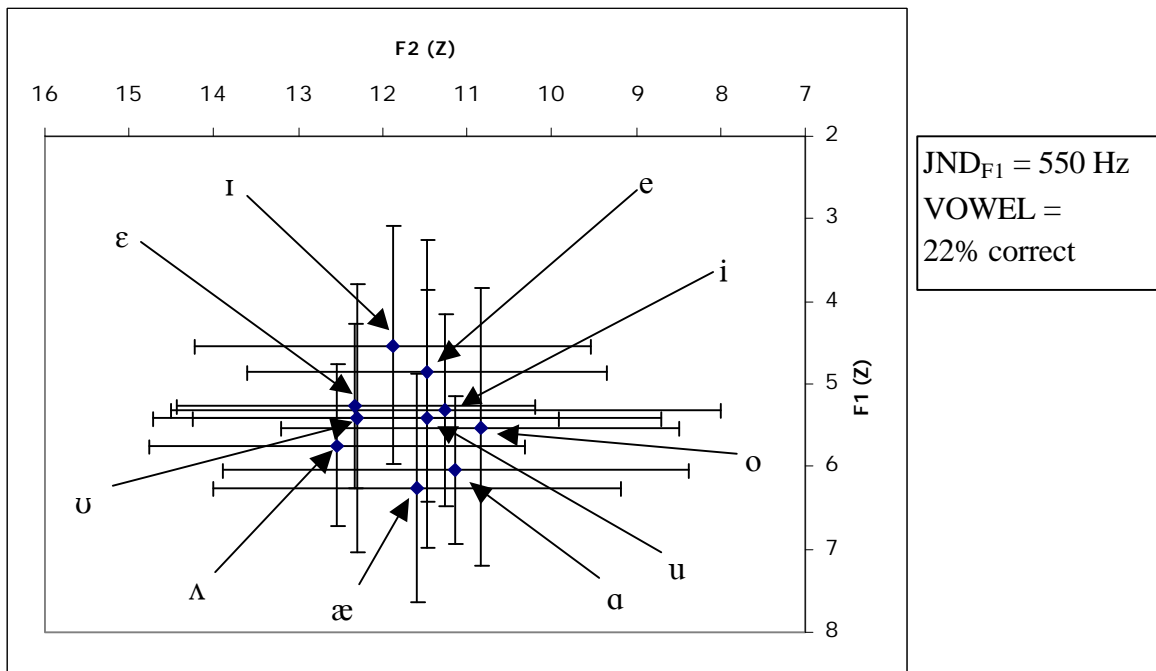


Figure 12: The vowel space of subject CI8.

The results indicate that most CI users adapted to their frequency shifted input, although they varied widely in the degree to which they adapted. In the vowel spaces, we find several that are composed of categories that overlap very little, and that appear to be in roughly the same regions in perceptual space as those obtained from normal hearing listeners. In contrast, the vowel spaces for CI7 and CI8 show much larger categories with a great deal of overlap between front and back vowels. The size of these categories and their arrangement in perceptual space have great difficulty in using spectral information to discriminate among vowel sounds, or in identifying particular vowels in running speech. The spaces of the CI users are smaller in size.

which gives the mean sizes of categories (in Bark) for each CI space, in the F1 and F2 dimension, along with the mean sizes averaged over both F1 and F2 across all CI users ("CI") and the equivalent scores for all normal-hearing subjects. The differences between the CI users and the normal-hearing subjects were in the order of a 3:1 ratio, and all were significant at the  $p \leq 0.01$  level (F1:  $t = 81.1$ ;  $df = 81.1$ ; F2:  $t = 6.9$ ,  $df = 6.9$ ; F1 and F2 combined:  $t = 163.6$ ,  $df = 163.6$ ). The vowel spaces of CI2 and CI3, the relatively "normal" spaces, have the smallest category sizes (along with CI1), while CI7 and CI8 have the highest. Large vowel categories typically overlap with one or more neighboring categories in perceptual space, which is reflected in the results of the F1 identification and vowel identification tests are consistent with this hypothesis: the subjects whose vowel identification result is reflected in Spearman rank correlations between the category size measures and the F1 identification scores correlated significantly ( $p \leq 0.05$ ), while the vowel identification test correlated with all three measures. All of the other correlations were only marginally significant ( $p \leq 0.05$ ).

Name	F1 (Z)	F2 (Z)	F1 and F2 (Z)
CI1	0.17	0.23	0.2
CI2	0.23	0.37	0.37
CI3	0.37	0.65	0.37
CI4	0.65	0.7	0.7
CI5	0.7	1.06	1.06
CI6	1.01	1.33	0.99
CI7	1.33	2.1	1.85
CI8	2.1	2.48	2.48
Mean CI	0.78		0.9
Mean NH		0.37	0.3

**Table 4:** The mean size of individual CI user's categories in the F1 dimension, the F2 dimension, and both dimensions, along with the mean category sizes across all normal-hearing (NH) and CI users' categories.

Test	Measure					
	F1		F2		F1 and F2	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
JND F1	0.78	0.04	0.63	0.1	0.72	0.06
VOWEL	-0.87	0.02	-0.93	0.01	-0.92	0.01

**Table 5:** Correlation between the category size measures and the F1 and vowel identification tests.

## Discussion

Despite the very large individual differences observed among cochlear implant subjects in all three tests used in this study, the results revealed an orderly relationship between the vowel spaces of these subjects and their scores on the F1 identification and the vowel identification tests. In the introduction we discussed two different potential limitations to vowel identification by cochlear implant users: limited frequency discrimination and limited ability to adapt to more-basal-than-normal stimulation. Overall, the major problem in vowel identification appears to be the former. Only one of the eight CI users, CI1, showed a systematic shift of her vowel space that seemed consistent with limited auditory plasticity. For each vowel, she selected regions with lower F1 and F2 than those selected by normal hearing listeners. This result is consistent with the proposal that she has not completely adapted to more-basal-than-normal stimulation and thus selects vowels with very low formants as the best exemplars, to compensate for the frequency shift that is imposed by the cochlear implant.

Other than CI1, none of the CI subjects showed any systematic shifts in their vowel spaces, suggesting that they were able to adapt to the frequency shift introduced by the cochlear implant. However, limitations in frequency discrimination played an important role in these subjects' ability to identify vowels. For example, CI7 and CI8 were the two CI users with the poorest  $JND_{F1}$  values: 350 Hz and 550 Hz respectively. Their vowel spaces showed substantial overlap among most vowel regions and their vowel identification scores were the lowest among the CI user group, 22% and 23% correct. Conversely, the CI users with the best (smallest)  $JND_{F1}$ 's, such as CI1 and CI2, tended to have high vowel identification scores and little overlap among vowel regions.

Our results are consistent with those of Hawks and Fourakis (1998). Their cochlear implant subjects also showed widely divergent amounts of overlap among vowel categories that were mapped using an identification task with synthetic stimuli. In terms of adaptation to more-basal-than-normal stimulation, the present results are encouraging because they suggest that most cochlear implant users are able to adapt to the shifts typically introduced by their devices. However, CI1's results remind us that not all listeners may be able to adapt in a similar way, and perhaps more listeners would find it difficult to adapt if the spectral shift was greater than that of the cochlear implant subjects in this study.

In conclusion, the data reported in this paper strongly suggest that vowel perception by cochlear implant users may be limited by the listener's formant frequency discrimination skills, in combination with

his/her ability to adapt to more basal- -normal stimulation. The present findings are also relevant to the over time) can provide much information about consonant identity. In future studies we intend to explore -basal than- to investigate the nature of the improvement in speech perception observed in most postlingually deaf applying the same methods used in this study to prelingually deafened CI users. In this case, we would not expect to see major prelingually deafened CI users may be, on average, even worse than that observed in postlingually deafened CI users. Finally, it may be interesting to determine whether the list have less neural and behavioral plasticity than the other listeners, or whether their lack of adaptation is due to a shallower insertion, which would result in a greater spectral shift to be overcome by the listener. Together, these studies will help us explain the enormous individual differences in speech perception by CI users, possibly paving the way for improved devices and intervention strategies in this clinical population.

## References

- Flanagan, J. (1957). E vowel sounds. *Journal of the Acoustical Society of America* 29, 533-534.
- Hawks, J.W. and Fourakis, M.S. (1998). Perception of synthetic vowels by cochlear implant recipients. *Journal of the Acoustical Society of America*, 104 (3) 1-5.
- Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). Acoustic characteristics of American vowels. *Journal of the Acoustical Society of America*, 98, 3099-3111.
- Wright, K., Flemming, E., and Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Journal of the Acoustical Society of America*, 94, 520-528.
- Kewley Port, D. and Watson, C. S. (1994). Formant structure of vowels. *Journal of the Acoustical Society of America* 95, 485-496.
- Klatt D. H., & Klatt L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 87, 820-830.
- Wright, K. (1999). Vowel spaces of dialects of English. In P. Eckert (Ed.), *New ways of analyzing sound change* (pp. 1-10). Cambridge, MA: MIT Press. *Journal of the Acoustical Society of America* 24, 175-184.
- Rosen, S., Faulkner, A., and Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts in speech. *Journal of the Acoustical Society of America*, 105, 3629-3636.
- Shannon, R.V., Fu, Q.-L., Chatterjee, M., Wygonski, J., Galvin, J., Zeng, F.-Wang, X. (1999). Speech processors for auditory prostheses, *Quarterly Progress Report #2* 10.

- Skinner, M.W., Clark, G.M., Whitford, L.A., Seligman, P.M., Staller, S.J., Shipp, D.B., Shallop, J.K., Everingham, C., Menapace, C.M., Arndt, P.L., Antogenelli, T., Brimacombe, J.A., Pijl, S., Daniels, P., George, C.R., McDermott, H., & Beiter, A.L. (1994). Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system. *American Journal of Otology*, 15 (Suppl. 2), 15-27.
- Skinner, M.W., Holden, L.K., and Holden, T.A. (1995). Effect of frequency boundary assignment on speech recognition with the SPEAK speech-coding strategy. *Annals of Otology, Rhinology, & Laryngology – Supplement*, 166, 307-11.
- Tyler, R. S., Preece, J. P., & Lowder, M. W. (1987). The Iowa audiovisual speech perception laser video disc. *Laser Videodisc and Laboratory Report*, University of Iowa at Iowa City, Department of Otolaryngology - Head and Neck Surgery.
- Wilson, B.S., Finley, C.C., Lawson, D.T., Wolford, R.D., Eddington, D.K. and Rabinowitz, W.M. (1991) Better speech recognition with cochlear implants. *Nature*, vol. 352, 18 July 1991, pp. 236-238.
- Wolfram, W. and Schilling-Estes, N. (1998). *American English*. Malden, MA: Blackwell Publishers.