Patterns of phoneme perception errors by listeners with cochlear implants as a function of overall speech perception ability

Benjamin Munson
Department of Communication Disorders, University of Minnesota, 115 Shevlin Hall,
164 Pillsbury Drive SE, Minneapolis, Minnesota 55455

Gail S. Donaldson and Shanna L. Allen[a]
Department of Otolaryngology, University of Minnesota, 8323 Philips Wangensteen Building,
516 Delaware Street SE, Minneapolis, Minnesota 55455

Elizabeth A. Collison[b]
Department of Communication Disorders, University of Minnesota 115 Shevlin Hall,
164 Pillsbury Drive SE, Minneapolis, Minnesota 55455

David A. Nelson
Department of Otolaryngology, University of Minnesota, 8323 Philips Wangensteen Building,
516 Delaware Street SE, Minneapolis, Minnesota 55455

(Received 22 March 2002; revised 14 November 2002; accepted 18 November 2002)

Many studies have noted great variability in speech perception ability among postlingually deafened adults with cochlear implants. This study examined phoneme misperceptions for 30 cochlear implant listeners using either the Nucleus-22 or Clarion version 1.2 device to examine whether listeners with better overall speech perception differed qualitatively from poorer listeners in their perception of vowel and consonant features. In the first analysis, simple regressions were used to predict the mean percent-correct scores for consonants and vowels for the better group of listeners from those of the poorer group. A strong relationship between the two groups was found for consonant identification, and a weak, nonsignificant relationship was found for vowel identification. In the second analysis, it was found that less information was transmitted for consonant and vowel features to the poorer listeners than to the better listeners; however, the pattern of information transmission was similar across groups. Taken together, results suggest that the performance difference between the two groups is primarily quantitative. The results underscore the importance of examining individuals' perception of individual phoneme features when attempting to relate speech perception to other predictor variables.
[DOI: 10.1121/1.1536630]

PACS numbers: 43.64.Me, 43.66.Ts, 43.66.Lj, 43.71.Ky, 43.71.Es [CWT]

I. INTRODUCTION

A large research literature has determined that cochlear implantation has the potential to greatly improve the speech perception of postlingually deafened adults. Postimplantation, improvements can be noted in phoneme perception scores, open-set word recognition, and sentence recognition. As implant technology has evolved, all of these measures have improved, with some listeners who utilize the most recent devices demonstrating speech perception in quiet approaching that of normal-hearing listeners [e.g., Wilson (2000)]. Nevertheless, there continues to be considerable variability in speech perception across individuals.

A large number of investigations have attempted to identify factors accounting for the variability in speech perception by individuals with cochlear implants. Studies have examined factors related to the implanted device [e.g., Tyler *et al.* (1996)], the speech processing strategy [e.g., Kompis *et al.* (1999); Osberger and Fisher (1999); Skinner *et al.* (1996, 1999)], listeners' psychophysical abilities [e.g., Busby and Clark (1999); Donaldson and Nelson (2000)], and higher-level cognitive and linguistic factors [e.g., Collison *et al.* (2002); Lyxell *et al.* (1998); Sarant *et al.* (1997)]. Although studies have related each of these factors to variability in speech perception performance, there is no consensus on their relative contributions to overall performance.

Most investigations of speech perception in individuals with cochlear implants have used percent-correct scores for vowel and consonant identification as the dependent measures of phoneme perception. These measures have the advantage of providing summary scores of phoneme perception, and appear to correlate with various factors purported to underlie speech perception. However, they have the disadvantage of collapsing performance across different phonemes, *potentially obscuring differences among listeners' error patterns*. Related to this, percent-correct scores may obscure qualitative differences between groups of listeners related to their use of particular acoustic and perceptual fea-

---

[a] Currently affiliated with the Henry Ford Hospital, Detroit, MI.
[b] Currently affiliated with the Indiana University—Purdue University of Indiana Medical Center, Indianapolis, IN.

tures. To address these deficits, a few studies have included analyses of consonant and vowel confusion matrices [e.g., Skinner *et al.* (1996, 1999)]. Confusion matrices catalogue the amount and types of misperceptions made by individual listeners or groups of listeners. They provide useful information regarding the specific phonemes that are most difficult for listeners to perceive, and common patterns of misperceptions. Additionally, confusion matrices can be analyzed using a variety of multivariate statistical techniques, such as log-linear modeling and multidimensional scaling, which provide information regarding the specific perceptual features that are most problematic for listeners.

Very few studies have compared patterns of phoneme errors in better- versus poorer-performing individuals with cochlear implants. Recently, one such study was reported by Van Wiringen and Wouters (1999). These investigators examined the phoneme perception errors of a group of individuals using the Laura cochlear implant. Listeners were divided arbitrarily into three groups based on their speech perception performance. Information transmission and multidimensional scaling analyses suggested that the three groups of listeners were using different features to perceive vowels and consonants, and that the better-performing listeners perceived features more efficiently. This study was limited, in that the high-performing listeners had only moderate speech perception performance: mean correct vowel and consonant identification for their best group of listeners was 66% and 49%, respectively. As a result, subgroups were not well separated on speech perception measures, and findings for the highest-performing group may not be applicable to the best users in current clinical populations.

The purpose of the present study is to further examine the differences between better-and poorer-performing cochlear implant users, by analyzing consonant and vowel misperceptions for two groups that were well separated in overall speech perception ability. A finding that the two groups differ qualitatively in their speech perception might suggest that different intervention strategies be used to enhance speech perception in the two groups. For example, a finding that poorer-performing listeners have more difficulty than better-performing listeners with vowel-height features but not vowel-backness features would imply that a remediation strategy should focus on the speech coding strategies targeting the first-formant frequency region. Analyses were designed to determine whether the speech perception errors made by poorer listeners were only quantitatively different than those made by better listeners (i.e., the poorer listeners made the same types of errors as better listeners, only more of them) or both quantitatively and qualitatively different (i.e., the poorer listeners made more errors than better listeners, and made errors on different features than the better listeners). In the first analysis, simple regression is used to compare better-and poorer-performing listeners' percent-correct scores for individual phonemes. The second analysis examines information transmission for vowel and consonant features to the two groups. Results from the two analyses suggest that the two groups differ primarily quantitatively in their perception of vowel and consonant sounds.

II. METHODS

A. Participants

Thirty postlingually deafened adults with cochlear implants participated. Summary data for the 30 participants are given in Table I. Twelve listeners were implanted with the Nucleus-22 device and used the SPEAK processing strategy. The other 18 listeners were implanted with the Clarion device; 13 of these listeners used the continuous interleaved sampling (CIS) strategy, four used the paired pulsatile (PPS) strategy, and one used the simultaneous analog (SAS) strategy. Listeners were heterogeneous with respect to many demographic variables, including etiology, age at onset, and duration of implant use.

B. Speech perception tests

Vowel and consonant recognition data were obtained using a standard phoneme-confusion procedure. Vowel stimuli were 11 /hVd/ monosyllables from the database of Hillenbrand *et al.* (1995), spoken by three male talkers. Only male talkers were used, because the female talkers in Hillenbrand *et al.*'s database demonstrated much greater variability in formant frequency than the male talkers, both within and between vowels. Vowels tested were /i, ɑ, ɛ, eɪ, ɝ, ɪ, æ, oʊ, ʊ, ʌ, u/ as in *heed*, *hod*, *head*, *hayed*, *heard*, *hid*, *had*, *hoed*, *hood*, *HUD* and *who'd*. Consonant stimuli were 19 /aCa/ disyllables from the stimulus set of Van Tasell *et al.* (1992), spoken by three male and three female talkers. Consonants tested were /p, t, k, b, d, g, f, θ, s, ʃ, v, ð, z, ʒ, m, n, r, l, j/.

Testing was conducted in a sound-isolated room, with the participant seated approximately 1 meter in front of a pair of high-quality loudspeakers and a video screen. Digitized speech stimuli were played out from computer memory, low-pass filtered at half the digitization rate, amplified, and presented through the speakers. The stimulus was presented once on each trial, and the subject used a computer mouse to select his or her response from a list of possible alternatives displayed on the video screen. The alternative choices presented in the vowel identification task were real words (e.g., *had*, *heed*), whereas alternatives for the consonant task were nonsense words (e.g., *awa, atha*). As in Van Tasell *et al.* (1992), correct-answer feedback was provided immediately after each stimulus presentation. Van Tasell *et al.* (1992) argued that it was appropriate to use feedback with examining performance of listeners who may not be well practiced in speech-perception tasks. The large number of different stimuli per block (six tokens for each of the 19 consonants or the 11 vowels) likely prevented subjects from remembering idiosyncratic features of individual tokens and identifying them based on those features. Stimulus level was calibrated such that average speech peaks of a file containing a concatenated sample of the speech perception stimuli reached 60 dB SPL on the slow, A scale of a Bruel & Kjær type 2203 sound-level meter at the location of the listener's head.

Vowel and consonant data were obtained in separate test sessions. For each stimulus type (vowels and consonants), one practice block and five standard blocks of data were obtained. Practice blocks were comprised of two trials per vowel phoneme (33 trials) or three trials per consonant pho-

TABLE I. Subject demographic information. Subjects ordered from highest mean consonant and vowel identification (top) to lowest (bottom).

Subject code	Age at onset (years)	Duration of profound deafness (years)	Duration of implant use (years; months)	NU-6 % words correct[a]	Etiology of deafness	Device-strategy	Group
N13	50	2	10;5	46	Progressive	Nucleus-SPEAK	Better
C04	36	0	0;3	66	Progressive	Clarion-CIS	Better
N29	3	28	1;7	90	Progressive	Nucleus-SPEAK	Better
N12	32	8	10;5	54	Progressive	Nucleus-SPEAK	Better
C15	33	7	1;0	68	Progressive	Clarion-CIS	Better
C12	34	13	0;10	···	Otosclerosis	Clarion-CIS	Better
C07	23	0	1;11	34	Progressive	Clarion-CIS	Better
N32	5	24	2;8	44	Rubella	Nucleus-SPEAK	Better
C14	16	47	1;6	76	Unknown	Clarion-CIS	Better
N14	49	0	6;9	88	Progressive	Nucleus-SPEAK	Better
C05	42	0	3;1	66	Unknown	Clarion-CIS	Better
C03	22	27	3;0	74	Progressive	Clarion-PPS	Better
C02	18	19	2;7	86	Unknown	Clarion-CIS	Better
C01	29	13	3;10	74	Progressive	Clarion-SAS	Better
N31	50	21	10;9	0	Noise	Nucleus-SPEAK	Poorer
N05	50	5	8;4	0	Otosclerosis	Nucleus-SPEAK	Poorer
C09	0	41	0;6	···	Unknown	Clarion-PPS	Poorer
N07	50	4	11;10	6	Cogan’s	Nucleus-SPEAK	Poorer
C13	72	6	1;2	50	Noise	Clarion-CIS	Poorer
N30	50	8	3;11	24	Otosclerosis	Nucleus-SPEAK	Poorer
C19	30	32	0;5	36	Progressive	Clarion-PPS	Poorer
C10	25	23	1;3	16	Meningitis	Clarion-PPS	Poorer
C06	50	12	2;1	24	Progressive	Clarion-CIS	Poorer
N09	56	0	11;0	24	Meniere’s	Nucleus-SPEAK	Poorer
C08	5	26	3;9	···	Ototoxicity	Clarion-CIS	Poorer
C20	28	31	1;3	48	Progressive	Clarion-CIS	Poorer
C17	23	17	0;10	32	Progressive	Clarion-CIS	Poorer
N22	66	1	6;3	30	Noise	Nucleus-SPEAK	Poorer
N28	57	0	4;9	58	Meningitis	Nucleus-SPEAK	Poorer
C11	43	0	0;10	···	Kearns-Sayre	Clarion-CIS	Poorer

[a] ··· indicates data not available.

neme (38 trials). Standard blocks were comprised of six trials per phoneme (66 vowels or 114 consonants) presented in random sequence. Occasionally, performance was observed to improve over the first few standard blocks. When this occurred, additional standard blocks were obtained until performance was stable for five consecutive blocks, and the final five blocks were retained. A merged confusion matrix was created from the five standard blocks of data for a particular subject. Each merged matrix represented 30 observations ($5\,\mathrm{blocks}\times 6\,\mathrm{tokens}$) per stimulus.

Subjects used their own speech processors for all testing. Nucleus subjects had a Spectra speech processor programmed in the SPEAK strategy. Clarion subjects had a v1.2 or S-series speech processor programmed in the CIS, PPS, or SAS strategy. Nucleus subjects adjusted the sensitivity controls on their processors to achieve comfortable loudness for the test stimuli. Clarion subjects adjusted the volume control, leaving the sensitivity control set to a level (“10:30” for v 1.2 users, “11:00” for S-series users) at which AGC compression would not be activated for the stimulus levels used.

The peak level chosen for presentation of stimuli in this study, 60 dB, is lower than that used in some other studies. A separate study [Donaldson and Smith (1999)] examined the effect of presentation level on perception of the same stimuli as were used in the current study. Because of the different ways that intensity is coded by the Nucleus and Clarion devices, presentation level may have impacted users of the two devices differentially. The Nucleus-22 SPEAK processor has only a sensitivity control, which adjusts the dynamic range of sounds encoded by the device upward and downward with respect to absolute sound levels. The dynamic range of the Nucleus device is approximately 30 dB: a given sensitivity setting might encode sounds between 30 and 60 dB SPL, whereas a higher sensitivity setting might encode sounds between 20 and 50 dB SPL. Sounds greater than those coded at the top of the dynamic range are all mapped to the top of the electric dynamic range, whereas sounds softer than the selected range are not coded at all. For testing, each Nucleus-22 subject adjusted the sensitivity of his/her processor so that the 60-dB phoneme stimuli were comfortably loud. The same sensitivity setting was used for the testing of consonants, vowels, and NU-6 words. Donaldson and Smith (1999) found that sound-field thresholds were quite variable across individuals utilizing the Nucleus-22 device. For the speech frequencies, they ranged from about 35 to 50 dB SPL, with the average thresholds for speech frequencies being 42.5 dB, and most occurring between 35 and 45 dB. Based on this observation, audibility may have been an issue for some cues; however, data on identification of consonants at different intensity levels did not support this hypothesis.

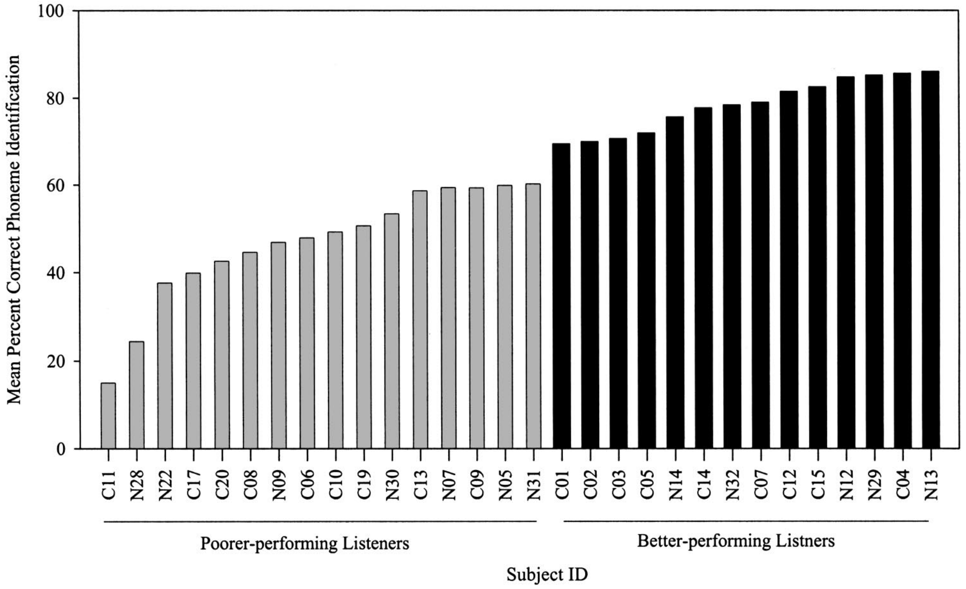

FIG. 1. Total percent-correct vowel and consonant identification for individual listeners.

For consonants, subjects showed small or absent improvements in performance when presentation level was increased from 60 to 70 dB SPL. When analyzed by feature, the average data showed no change between 60 and 70 dB for voicing and place features, but did show about 10% improvement in transmitted information for manner features. For vowels, subjects showed fairly constant performance between 50 and 70 dB, so audibility would not seem to be an issue at 60 dB.

In contrast, the Clarion device has both a sensitivity control and a volume control. Subjects set their sensitivity to the default setting (“10:30” or “11:30” depending on processor type), which sets the top of the encoded dynamic range to approximately 70 dB SPL. They then adjusted the volume control to achieve comfortable loudness for 60-dB stimuli. The volume control for this device moves the top of the *electric* dynamic range up or down to alter loudness. For the Clarion device, the range of speech sounds encoded (referred to as *input dynamic range* [IDR]) is adjustable for values between 30 and 60 dB. Almost all subjects used an IDR of 50 or 60 dB. As would be expected from this wider input dynamic range, Clarion subjects had more sensitive sound-field thresholds than Nucleus subjects, average 28 dB SPL for the speech frequencies. Clarion subjects reached asymptotic performance for consonants by 60 dB and for vowels by 50 dB. For many individual subjects, asymptotic performance was reached for a lower presentation level.

In summary, data from Donaldson and Smith (1999) suggest that audibility could be a minor factor influencing consonant performance in Nucleus subjects, but not for vowel performance in Nucleus subjects, and not for consonant or vowel performance in Clarion subjects.

III. RESULTS

A. Descriptive data

1. Percent-correct scores

The analyses presented in this section compare the performance of better- and poorer-performing listeners with cochlear implants. For the purposes of this study, we defined better- and poorer-performing listeners using the following technique. First, we calculated total percent-correct identification for consonants and vowels. Second, we examined individual listeners' scores on this composite measure. In doing so, we observed that no listeners had composite scores between 60% and 69%. This is illustrated in Fig. 1.

We divided the larger group of listeners into two subgroups. The better-performing group of listeners ($n=14$) had a mean composite score of 78.5% (s.d.$=6.1$, range $=[69\%-86\%]$). The poorer-performing group of listeners ($n=16$) had a mean composite score of 46.9% (s.d.$=13$, range$=[16\%-60\%]$). The two groups were well separated in their percentage-correct scores for vowels ($M=86.6\%$, s.d.$=5.8$ for the better listeners, $M=53.7\%$, s.d.$=16$ for the poorer listeners) and consonants ($M=70.4\%$, s.d.$=7.8$ for the better listeners, $M=40\%$, s.d.$=12.8$ for the poorer listeners). Nucleus-22 and Clarion users were equally distributed across the two groups, as indicated by a chi-square test ($\chi^2=0.36, p>0.05$). Mann-Whitney U tests indicated that the two groups did not differ in their age at onset of deafness, duration of profound deafness, or duration of implant use. Percent-correct scores for identification of NU-6 words in quiet were available for 13 of the poorer listeners and 13 of the better listeners. Mann-Whitney U tests confirmed that the two groups differed in their word perception ($z=-3.774$, $p<0.001$). The better listeners had a mean percent-correct score of 67% (s.d.$=18\%$). The poorer-performing group had a mean score of 27% (s.d.$=18\%$). Each listener's group membership is noted in Table I.

2. Common patterns of phoneme confusions

Prior to completing statistical analyses, we conducted a descriptive analysis of the most common patterns of phoneme confusions made by the two groups of listeners. The confusion matrices for vowels and consonants by the poorer- and better-performing groups are presented in Tables II–V.

Phoneme confusions are reported in the form response/stimulus (i.e., t/k means *the stimulus* [k] *was misperceived to*

TABLE II. Vowel confusion matrix, poorer listeners.

	æ	ɑ	ɛ	eɪ	ɝ	ɪ	i	oʊ	ʊ	ʊ	u
æ	303	28	66	20	30	16	12	4	2	4	1
ɑ	143	207	11	6	61	2	1	22	6	22	5
ɛ	19	5	186	14	4	186	11	4	25	30	2
eɪ	21	6	16	305	12	17	98	3	2	2	4
ɝ	8	11	29	8	240	13	2	36	47	27	65
ɪ	6	0	66	12	3	356	30	1	4	7	1
i	10	0	11	49	6	18	387	4	1	0	0
oʊ	2	17	7	4	33	2	2	231	31	8	149
ʊ	7	10	28	4	22	39	4	31	257	47	37
ʌ	11	19	64	3	19	38	6	15	164	134	13
u	10	12	7	7	54	5	1	94	38	12	246

be [*t*]). The ten most common vowel confusions for the poorer-performing group of listeners were æ/ɑ, ɛ/ɪ, ɛ/ʌ, ɝ/ɑ, ɪ/ɛ, i/eɪ, oʊ/u, ʊ/ʌ, u/ɝ, and u/oʊ. These confusions suggest problems perceiving a variety of vowel features. The pairs i/eɪ and oʊ/u suggest problems with height; æ/ɑ, ɛ/ɪ, ɛ/ʌ, and, ɪ/ɛ suggest problems with backness, and ɝ/ɑ and u/ɝ indicate a difficulty perceiving the r-coloring feature of the vowel /ɝ/. The ten most common misperceptions made by the better-performing group of listeners were æ/ɑ, ɛ/ɪ, ɛ/æ, ɝ/u, ɪ/ɛ, oʊ/u, ʊ/ɝ, ʊ/ʌ, u/ʌ, and u/ʊ. Five of the most frequent misperceptions were common to the two groups. Again, these confusions show a problem perceiving a variety of vowel features.

The ten most common consonant confusions for the poorer-performing group of listeners were t/p, k/p, t/k, g/d, f/θ, f/s, ʃ/s, v/ð, m/n, and l/r. These confusions indicate particular difficulty with obstruent sounds: eight of the ten target stimuli were stops or fricatives, six of which were voiceless. The ten most common consonant confusions for the better-performing group of listeners were t/p, t/k, g/d, f/θ, θ/f, θ/s, v/ð, ð/v, ð/z, and m/n. Again, these errors show a particular difficulty perceiving obstruent sounds, as nine of the targets were stops or fricatives. Unlike the poorer group of listeners, the target obstruent sounds most often misperceived by the better-performing listeners were equally likely to be voiced and voiceless; the poorer-performing listeners were most likely to make errors perceiving voiceless obstruents. In general, both groups showed similar patterns of confusion for consonants. Six of the most common misperceptions were evidenced by both of the two groups. In addition, two of better-performing users' most common confusions, θ/f and θ/s, were among the poorer-performing listeners' most common misperceptions, albeit not among the ten most common. Only two of the better-performing listeners' errors, ð/v and ð/z, were not among the poorer-performing listeners' most common misperceptions. This is not surprising; given that [ð] was the poorer-performing listeners' least-common response to the set of stimuli.

In general, the commonality in error vowel and consonant error patterns between the two groups suggests that perception of these two classes of sounds differs quantitatively between the two groups. Both groups demonstrate difficulty in perceiving place of articulation features in obstruent consonants, and a variety of vowel features.

B. Regression analysis of percent-correct scores

In the first statistical analysis, we computed mean percent-correct scores for each phoneme for both the better- and poorer-performing groups of listeners, and measured the degree of association between groups using simple regression analyses. We reasoned that the strength of the association would be a measure of the extent to which two groups' error patterns were qualitatively similar. If the poorer-performing users' errors differ only quantitatively from the better-performing users', then we would expect a close-to-perfect measure of association (i.e., an R^2 approaching 1.0). In contrast, if the performance of the poorer-performing group of listeners were to some degree qualitatively different from the better-performing listeners', then a less-than-perfect measure of association would be expected. In addition, we examined the pattern of outlying residuals in the regression

TABLE III. Vowel confusion matrix, better listeners.

	æ	ɑ	ɛ	eɪ	ɝ	ɪ	i	oʊ	ʊ	ʌ	u
æ	358	4	26	2	0	0	0	0	0	0	0
ɑ	25	357	1	0	1	0	0	1	0	4	1
ɛ	8	1	310	1	0	55	1	0	3	11	0
eɪ	0	0	1	374	3	2	10	0	0	0	0
ɝ	0	0	3	0	342	3	1	3	22	3	13
ɪ	0	0	48	4	2	332	3	0	1	0	0
i	0	0	3	14	0	0	371	2	0	0	0
oʊ	0	0	0	0	0	1	1	328	1	3	56
ʊ	1	1	6	1	9	1	1	7	320	9	34
ʌ	0	2	10	0	2	0	0	1	52	322	1
u	0	0	0	1	22	0	0	60	8	4	295

TABLE IV. Consonant confusion matrix, poorer listeners.

	p	t	k	b	d	ɡ	f	θ	s	ʃ	v	ð	z	ʒ	m
p	188	98	106	12	6	15	17	13	5	1	3	3	2	2	0
t	55	234	135	2	2	5	8	22	3	4	1	3	2	2	0
k	72	115	242	1	1	4	7	11	9	4	0	5	1	0	1
b	32	3	6	190	55	60	31	21	3	0	33	10	3	0	16
d	13	12	9	42	151	200	9	11	0	0	7	7	0	0	4
ɡ	14	16	15	15	80	268	7	9	3	1	11	5	6	2	2
f	39	14	13	11	1	4	177	86	55	17	20	13	5	8	3
θ	44	16	12	9	8	4	171	127	50	5	5	19	4	0	0
s	24	8	6	7	7	4	102	79	128	46	10	21	11	18	1
ʃ	2	5	0	0	1	0	5	1	49	353	1	2	11	46	1
v	15	9	8	55	16	39	16	42	19	2	101	22	25	6	26
ð	10	7	6	66	24	28	22	51	7	2	101	23	21	8	17
z	9	2	5	22	11	14	17	39	31	32	57	9	79	43	5
ʒ	1	4	2	1	0	3	4	3	7	103	5	2	54	258	2
m	7	0	7	23	2	5	19	4	4	1	14	2	6	2	217
n	7	6	6	8	21	12	4	4	5	1	16	2	12	2	67
r	4	0	2	14	2	7	5	6	5	0	56	5	10	5	22
l	5	1	0	17	1	4	5	1	3	1	12	2	5	0	78
j	2	5	5	1	4	10	4	3	6	12	11	2	24	13	6

analysis. Our rationale for this analysis was that the phonemes perceived differently by the two groups would be associated with the largest residuals. A similar technique has been used previously to examine differences in reaction times between children with slower and faster processing of linguistic and nonlinguistic stimuli [Lahey *et al.* (1999)].

Vowels and consonants were examined separately. In both analyses, the mean scores for phoneme identification for the poorer-performing group served as the dependent variable in the regression, with the mean scores for the better-performing group serving as the independent variable. For consonants, a strong, significant association was found between scores for the two groups of listeners [$F(1,17) = 117.823$, $p<0.001$, $R^2=0.874$]. The relationship between the two groups is shown in Fig. 2. We chose to examine phonemes whose standardized residuals were greater than 1.0 or less than -1.0, indicating sounds that fell outside of the 68% confidence interval of the regression line. When the standardized residuals were examined, it was found that the consonants /ʃ/ and /f/ had standardized residuals greater than 1.0 ($z=2.23$ for /ʃ/, $z=1.19$ for /f/). Thus, the poorer-performing listeners were identifying the consonants /ʃ/ and /f/ with greater accuracy than would be predicted by the performance of the better-performing listeners: mean correct identification for /ʃ/ was 74% for the poorer-performing group of listeners and 93% for the better-performing group of listeners; /f/ was identified with 37% accuracy for the poorer-performing group of listeners and 58% by the better-performing group of listeners. In contrast, the consonants /l/, /n/, /r/, and /p/ were found to have standardized residuals less than -1 ($z=-1.05$ for /n/, $z=-1.24$ for /r/, $z=-1.3$ for /l/, $z=-1.73$ for /p/). That is, the poorer-performing listeners identified the consonants /l/, /n/, /r/, and /p/ with lower accuracy than would be predicted by the performance of the better-performing group of listeners. Mean percent-correct scores for the two groups were as follows: for /l/, $M=27\%$

TABLE V. Consonant confusion matrix, better listeners.

	p	t	k	b	d	ɡ	f	θ	s	ʃ	v	ð	z	ʒ	m
p	342	50	13	1	0	1	3	2	1	1	0	0	0	0	0
t	30	311	64	0	0	0	0	3	0	6	0	0	0	0	0
k	25	52	329	0	0	1	2	2	0	0	0	1	0	0	0
b	0	0	0	300	50	7	11	12	0	0	15	11	1	0	4
d	0	0	0	7	238	160	1	0	0	0	0	6	2	0	0
ɡ	0	2	2	5	27	371	2	2	1	0	0	2	2	0	0
f	5	2	0	7	1	2	240	60	61	7	15	8	3	0	2
θ	2	0	0	7	2	1	135	198	49	0	8	12	1	0	1
s	1	1	1	2	0	3	49	94	237	6	3	11	7	0	0
ʃ	0	3	0	0	0	0	0	0	18	385	0	0	1	8	0
v	8	3	0	47	8	0	11	13	1	1	187	61	29	0	16
ð	0	0	0	40	21	18	2	32	0	0	98	130	39	0	6
z	1	0	0	1	12	13	8	16	27	11	15	60	202	13	0
ʒ	0	1	0	0	0	0	0	0	2	15	0	0	23	364	0
m	1	0	0	2	0	0	0	2	0	0	2	2	0	0	315
n	1	0	0	0	3	0	0	0	0	0	1	3	0	0	19
r	0	0	0	0	0	0	0	3	0	0	11	1	1	0	0
l	1	0	2	0	0	0	0	0	0	0	4	0	1	0	19
j	0	0	0	0	0	0	0	0	0	0	3	0	3	0	0

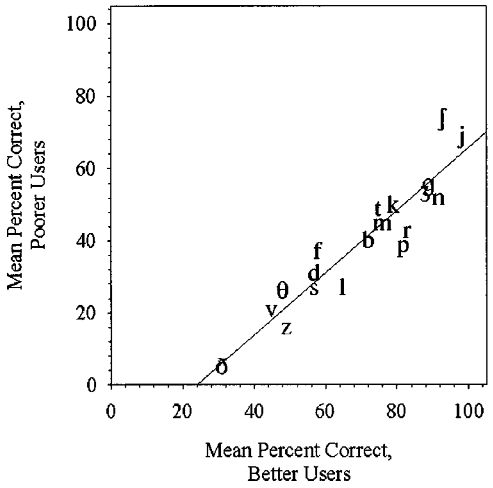

FIG. 2. Mean percent of consonants correctly perceived by the better listeners (x axis) and the poorer listeners (y axis).

for the poorer group and $M=65\%$ for the better-performing group; for /n/ $M=52\%$ for the poorer-performing group and $M=92\%$ for the better-performing group; for /r/, $M=43\%$ for the poorer-performing group and $M=83\%$ for the better-performing group; for /p/, $M=39\%$ for the poorer-performing group and $M=82\%$ for the better-performing group.

When vowels were examined, a weak, statistically nonsignificant relationship between the two groups was found ($F[1,9]=3.213$, $p=0.107$, $R^2=0.263$). The relationship between the two groups is shown in Fig. 3. Residuals were not examined for vowels: because the slope of the regression line was not significantly different from zero, the magnitude of the standardized residuals could not be considered stable estimates.

A possible explanation for the difference in regression results for consonants and vowels is that a greater range of performance was noted in consonants than in vowels. For consonants, poorer-performing listeners had a range of 69% in their percent-correct scores, from 5% for the phoneme /ð/ to 74% for the phoneme /ʃ/. Better-performing listeners had a range of 68%, from 31% for the phoneme /ð/ to 98% for the phoneme /j/. Performance on vowels was less variable. For the poorer listeners, performance ranged 53%, from 27% for /ʌ/ to 80% for the vowel /i/. Performance of the better listeners ranged only 19%, from 77% for the vowel /u/ to 96% for the vowel /eɪ/. In addition, fewer vowels than consonants were examined, suggesting that the lack of significance in the regression on vowel performance may have been due to statistical power.

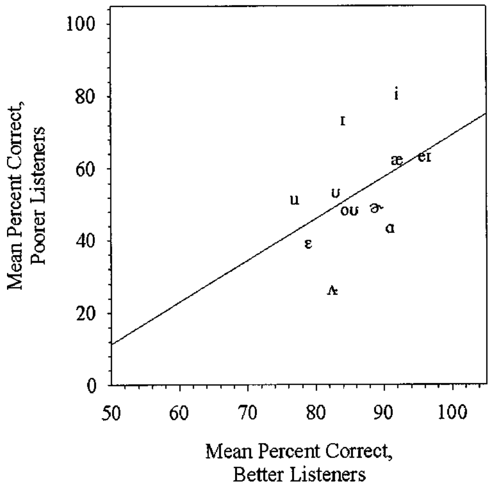

FIG. 3. Mean percent of vowels correctly perceived by the better listeners (x axis) and the poorer listeners (y axis).

To examine whether these factors influenced the different regression results for vowels and consonants, we ran two additional regressions on subsets of the consonant performance data. First, we ranked the consonants based on the percent-correct scores of the poorer-performing listeners. Based on this, we divided the consonants into two subsets of nine consonants: poorly perceived consonants (/z/, /v/, /θ/, /s/, /l/, /d/, /f/, /p/, and /b/) and better-perceived consonants (/r/, /m/, /t/, /k/, /n/, /ʒ/, /g/, /j/, /ʃ/). The consonant demonstrating the poorest perception by both groups, /ð/, was omitted from this analysis so that the two subsets of consonants could be equal in size. The range of performance for the poorly perceived consonants was 24% for the poorer listeners (from 16% for /z/ to 40% for /b/) and 37% for the better listeners (from 45% for /v/ to 82% for /p/). The range of performance for the better-perceived consonants was 31% for the poorer listeners (from 43% for /r/ to 74% for /s/) and 33% for the better listeners (from 75% for /t/ to 98% for /j/). Separate regressions were run on these two subsets of consonants. These separate regressions were more comparable to the regression analyses of vowel data, both in terms of range of performance and number of items. Even with fewer items and a smaller range of performance, both of the regressions were significant [$F(1,7)=11.962$, $p=0.011$, $R^2=0.63$ for the poorly perceived consonants; $F(1,7)=10.363$, $p=0.015$, $R^2=0.60$ for the better-perceived consonants]. These analyses suggest that the difference in vowel and consonant regression results between the two groups may not be due to differences in range of performance or number of items. However, this conclusion is limited by the fact that the smallest range of performance on the consonant subsets (24% for the poorest-perceived consonants by the poorer-performing listeners) was still larger than the 19% range in the better-performing listeners' vowel perception.

A second set of analyses compared the percent-correct data for individual poorer-performing listeners to the mean data for the better-performing group of listeners. Each of the 16 poorer-performing users' percent correct vowel and consonant identification served as the dependent variable in a series of regressions, with the mean scores for the better-performing group of listeners serving as the independent variable. The results of these 32 regressions, including R^2 and p values, supported the group analysis: the majority of poorer-performing listeners ($n=13$) showed significant relationships for consonants, but not for vowels. In contrast to the descriptive analysis of confusion matrices, these results suggest that the two groups differ qualitatively and quantitatively in the perception of vowels, but not in the perception of consonants. Again, this conclusion is limited by the fact that larger ranges in performance in the better-performing

TABLE VI. Features used for vowels and consonants in the information transmission analysis.

Vowels			Consonants		
Feature	Categories	Items	Feature	Categories	Items
Tense/lax	tense	/i,e,æ,ɑ,o,u,/	voicing	voiced	/b,d,g,v,ð,z, ʒ,m,n,j,r,l/
	lax	/ɪ,ɛ,ʊ,ʌ/		voiceless	/p,t,k,f,θ,s,S/
$f0$	low	/i,e,æ,ɑ/	duration	short	/d,g,k,b,p,t,j/
	mid	/ɪ,u,ɛ/		medium	/l,n,m,v,ð,r/
	high	/o,ʊ,ɚ,ʌ/		long	/ʒ,z,f,ʃ,θ,s/
Height	low	/æ,ɑ/	place	labial	/p,b,m,f,v/
	mid	/e,ɛ,ʌ,ɚ,o/		coronal	/θ,ð,t,d,n,l,r/
	high	/i,ɪ,u,ʊ/		palatal	/ʃ,ʒ,j/
Backness	back	/o,ʊ,u/		velar	/k,g/
	central	/ɑ,ʌ,ɚ/	manner	plosive	/p,b,t,d,k,g/
	front	/i,ɪ,e,ɛ,æ/		fricative	/f,v,θ,ð,s,z,ʃ,ʒ/
r-coloring	r-colored	/ɚ/		sonorant	/m,n,j,r,l/
	non-r-colored	/i,ɪ,e,ɛ,æ,ɑ,o, ʊ,u,ʌ/			

listeners were noted for consonant perception than for vowel perception.

C. Information transmission for better- and poorer-performing listeners

The second statistical analysis used sequential information transmission analysis [SINFA, Wang and Bilger (1973)] to examine whether better- and poorer-performing users' vowel and consonant perception differed for individual features. We reasoned that if the two groups' perception of vowels and consonants differed qualitatively, an analysis of variance (ANOVA) would reveal a significant interaction between group and feature.

The features used for the analysis of consonants and vowels are listed in Table VI. The features used for consonants were voicing, place of articulation, manner of articulation, and duration. The first three of these features are traditionally used in describing consonants [e.g., Ladefoged (2001)]. The feature duration was included in case some of the poorest listeners were unable to use enough spectral and envelope cues to receive any information about voicing, place, or manner, and were able to perceive only the duration of consonants. Values for the duration feature were determined by measuring the mean durations of the consonantal portion of the /ɑCɑ/ stimuli and categorizing them as short (mean duration of 100 to 110 ms), medium length (mean duration of 123 to 136 ms), or long (mean duration of 152 to 222 ms). These categories were chosen arbitrarily, based on the observation that there were no consonants with mean values between 110 and 123 ms, or between 136 and 152 ms. The features height, backness, r-coloring, tense/lax, and $f0$ were used for vowels. For the feature $f0$, mean $f0$'s were measured for the stimuli, and vowels were classified as either low pitch (mean $f0$ of 104 to 112 Hz), medium pitch (mean $f0$ of 118 to 122 Hz), or high pitch (mean $f0$ of 131 to 152 Hz).

SINFA was used to calculate the information transmitted by each feature to each listener. Information transmission scores for individual listeners were then submitted to a two-way mixed-model ANOVA. Vowels and consonants were analyzed separately. For consonants, feature (voicing, place, manner, and duration) was the within-subjects factor and group (better versus poorer listeners) was the between-subjects factor. Mean information transmitted for the four features to listeners in the two groups can be found in Fig. 4. A significant main effect of group was found [$F(1,28) = 38.308$, $p < 0.001$], with less information transmitted to the poorer listeners than to the better listeners. A significant main effect of feature was also found [$F(3,84) = 53.455$, $p < 0.001$]. All *post hoc* differences were significant, with

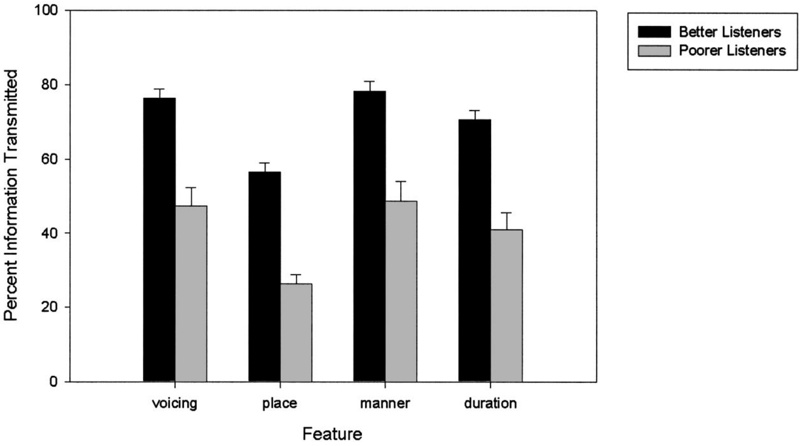

FIG. 4. Percentage of information transmitted for consonant features. Error bars represent one standard error of measurement.

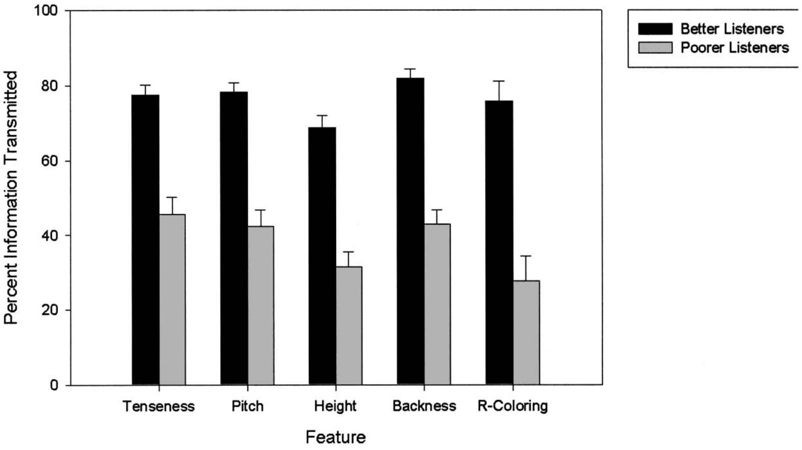

FIG. 5. Percentage of information transmitted for vowel features. Error bars represent one standard error of measurement.

the exception of that between voicing and manner of articulation. No significant group by feature interaction was found [$F(3,84)=0.049$, $p=0.985$], suggesting that the two groups did not differ qualitatively in their perception of consonant features.

The second ANOVA examined information transmitted for vowels. Again, information transmission scores were submitted to a two-way mixed model ANOVA, with feature (tense/lax, $f0$, height, backness, and r-coloring) as the within-subjects factor and group (poorer versus better) as the between-subjects factor. Mean information transmitted for the five features to the listeners in the two groups can be found in Fig. 5. A significant main effect of group was found [$F(1,28)=60.414$, $p<0.001$]. Less information was transmitted to the poorer-performing listeners than to the better-performing listeners. A significant main effect of feature was also found [$F(4,112)=8.796$, $p=0.001$]. Less information about the feature height (corresponding to $F1$ frequency) was transmitted than the other features. The group by feature interaction was nearly significant [$F(4,112)=2.353$, $p=0.057$]. Visual inspection of the data suggests that the group difference between better- and poorer-listeners was greater for the features height ($F1$ frequency) and r-coloring ($F3$ frequency) than for the other three features, with large within-group variability noted for the r-coloring feature. Although this interaction was not significant, the trend is consistent with the findings of the regression analyses, suggesting that the two groups may differ qualitatively in their perception of vowels.

The participants of this study included both Nucleus-22 and Clarion users. As noted earlier, there were no significant differences in the number of Clarion and Nucleus-22 users in the better- and poorer-performing groups of listeners. However, given that these two devices and their associated speech-processing strategies code sound very differently, we were interested in examining whether information transmitted would be different for the two groups. Two mixed-model ANOVAs examined the influence of group membership (the between-subjects factor) and feature (the within-subjects factor) on information transmitted. No significant effect of group was found for either analysis. In addition, no group-by-feature interactions were found, despite the potential small difference in audibility of consonant features by users of the Nucleus-22 and Clarion devices.

The consonant results are similar to those in previous studies, some of which used a different set of consonant stimuli and consonant features [Fu and Shannon (1999); McKay and McDermott (1993); Skinner *et al.* (1996, 1999); Tyler and Moore (1992); Van Wiringen and Wouters (1999); Vandali (2001)]. In all of these studies, place of articulation information was more poorly transmitted than information about manner and voicing. There is less previous research concerning information transmission in vowel perception to listeners with cochlear implants. A previous study comparing vowel feature perception for Nucleus users utilizing SPEAK speech-processing strategies was presented by Skinner *et al.* (1999). Consistent with the current study, Skinner *et al.* (1999) found that more information was transmitted for backness ($F2$) features than for height or r-coloring features. An earlier study examined information transmission in vowels for listeners with an Ineraid cochlear implant [Tyler *et al.* (1992)], and found that the greatest amount of information was transmitted for duration and $F1$ frequency. This result contrasts with the results for the current study, in which the least information was transmitted for the $F1$ feature. This may be attributable to differences between the devices and speech processing strategies used in that study and in the current study.

The results of the information-transmission analysis suggested that the two groups of listeners did not differ qualitatively in their perception of either vowel or consonant features. No group-by-feature interactions were found, although a nonsignificant trend was noted for poorer perception of height and r-coloring features by the poorer-performing group of listeners.

IV. DISCUSSION

This study examined patterns of phoneme misperceptions by two groups of cochlear implant listeners, those with relatively better speech perception performance, and those with relatively poorer performance. The three analyses performed in this paper gave conflicting findings. Regression analyses of the two groups' percent-correct vowel and consonant identification scores showed a strong, significant relationship for consonant perception and a weak, nonsignificant relationship for vowel perception. This suggests that poorer-performing listeners' consonant identification differed only quantitatively from better-performing listeners', while the vowel perception differed both quantitatively and qualitatively. A larger range of performance was noted in consonant perception than in vowel perception for the two groups. In addition, more consonants were included in the analysis than vowels. Thus, the apparent qualitative difference between the two groups' vowel perception may have been due in part to the restriction-of-range and ceiling effects in the better-performing group of listeners.

In contrast to the regression findings, descriptive analysis of confusion matrices found similar patterns of confusion for both vowels and consonants, and information transmission analysis suggested that the same relative amount of information of vowel and consonant features was being transmitted to listeners in the two groups. Results of the information transmission analysis suggested that the poorer-performing listeners were receiving less information about height (which is correlated with mean $F1$ values) and r-coloring (which is correlated with mean $F3$ values) than $F0$, $F2$, and duration. Taken together, these findings suggest that the poorer-performing listeners had greater difficulty perceiving spectral information in the frequency region of $F1$ and $F3$. The better-performing listeners were able to use spectral information from three distinct spectral bands, $F1$, $F2$, and $F3$, while the poorer-performing listeners were able, in general, to hear differences best in $F2$. This finding may suggest that the poorer-performing listeners would benefit from a speech-processing strategy in which more electrodes are dedicated to representation of $F1$ and $F3$ information. This is finding contrasts with those of Henry *et al.* (2000) and McKay and Henshall (2002), who found that low-frequency information was best transmitted to the poorest-performing listeners with Nucleus-22 cochlear implants.

In general, the information suggesting that the two groups did not differ qualitatively on either vowel or consonant perception was stronger than evidence to the contrary. This finding is important, because it suggests that techniques to enhance the speech perception of people who use cochlear implants need not take into account overall level of functioning. The results of this study expand on the results of Van Wiringen and Wouters's (1999) study of the speech perception of better-, intermediate-, and poorer-performing individuals with cochlear implants. Van Wiringen and Wouters (1999) found that better-, intermediate-, and poorer-performing listeners used different features to perceive consonants, while the two groups of listeners in the current study appeared to use the same features. The differences between that study and the current one may be due to differences in the overall levels of performance in the two studies; to differences the different devices used by the listeners; to differences in the homogeneity of performance within the different groups; and to differences in the languages being examined.

In contrast to the regression analyses, the SINFA analysis found that the two groups used the same features to perceive both vowels and consonants. One problematic aspect of SINFA analysis is that it assumes that articulatory features are relevant for speech perception. While early theories of speech perception [e.g., Liberman and Mattingly (1985)] emphasized the possible articulatory basis of speech perception, more recent theories have emphasized the importance of acoustic information and auditory processing [e.g., Kluender and Lotto (1999)]. Consonant features such as [place] and [voice] are much more uniform articulatorily than acoustically. For example, the feature [place] has different acoustic correlates depending on the manner of articulation of the consonant being produced. Measures of the reception of the feature [place] in consonants provides limited information regarding the specific acoustic parameters that were used, and thus provides relatively little information on which parameters should be enhanced in individuals demonstrating poor reception of place features. Vowel features used in SINFA analyses are much more transparently related to acoustic features. For example, the feature [height] is well correlated with $F1$ frequency, and the feature [back] is well correlated with $F2$ frequency. The results of the information-transmission analysis might have been affected by the different relationships between articulation and acoustics for consonants and vowels.

In summary, these results provide tentative support for the hypothesis that listeners varying in overall performance do not differ qualitatively in their consonant and vowel perception. The fact that not all analyses arrived at this conclusion underscores the methodological difficulties of assessing the perception of specific vowel and consonant features with tests of vowel and consonant identification accuracy. Researchers wishing to understand the features that are misperceived by listeners with cochlear implants must make inferences based on patterns of consonant and vowel identification in these tasks. As an alternative to this, researchers could measure speech perception by examining the identification and discrimination of a small number of speech contrasts whose perception depends on a single acoustic parameter, rather than with mean data on accuracy of identification of a large set of phonemes.

ACKNOWLEDGMENTS

We gratefully acknowledge James Hillenbrand for providing the vowel identification stimuli, and for sharing the formant frequency and duration measures for the vowels used in the speech perception experiments. We thank Dianne Van Tasell for providing the stimuli used in the consonant identification experiment. We thank Peggy Nelson, Arlene Carney, the two anonymous reviewers, and the editor for valuable comments on this work. We also thank Edward Carney for technical assistance. This project was supported by NIDCD grant number P01-DC00110 and by the Lions 5M International Hearing Foundation.

Bosman, A. (**1996**). "Confusion analysis in the assessment of speech perception and hearing aids," in *Psychoacoustics: Speech and Hearing Aids*, edited by B. Kollmeier (World Scientific, Singapore).

Busby, P. A., and Clark, G. M. (**1999**). "Gap detection by early-deafened cochlear-implant subjects," J. Acoust. Soc. Am. **105**, 1841–1852.

Collison, E., Munson, B., and Carney, A. (**2002**). "Relationships among vocabulary size, nonverbal cognition, and spoken word recognition in adults with cochlear implants," J. Acoust. Soc. Am. **111**, 2428(A).

Donaldson, G. S., and Nelson, D. A. (**2000**). "Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies," J. Acoust. Soc. Am. **107**, 1645–1658.

Donaldson, G. S., and Smith, S. L. (**1999**). ''Speech performance-intensity functions for Nucleus SPEAK and Clarion CIS cochlear implant listeners,'' American Academy of Audiology Annual Convention, Miami Beach, FL.

Fu, Q.-J., and Shannon, R. (**1999**). ''Effects of electrode location and spacing on phoneme recognition with the Nucleus-22 cochlear implant,'' Ear Hear. **20**, 321–331.

Henry, B., McKay, C., McDermott, H., and Graeme, C. (**2000**). ''The relationship between speech perception and electrode discrimination in cochlear implantees,'' J. Acoust. Soc. Am. **108**, 1269–1280.

Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K., (**1995**). ''Acoustic characteristics of American English Vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Kluender, K., and Lotto, A. (**1999**). ''Virtues and perils of an empiricist approach to speech perception,'' J. Acoust. Soc. Am. **105**, 503–511.

Kompis, M., Vischer, M. W., and Hausler, R. (**1999**). ''Performance of compressed analogue (CA) and continuous interleaved sampling (CIS) coding strategies for cochlear implants in quiet and noise,'' Acta Oto-Laryngol. **119**, 659–664.

Ladefoged, P. (**2001**). *A Course in Phonetics*, 4th ed. (Harcourt College Publishers, New York).

Lahey, M., Edwards, J., and Munson, B. (**1999**). ''SLI: General or Specific Processing Limitations?'' Miniseminar presented at the 1999 Annual Convention of the American Speech-Language-Hearing Association, San Francisco, CA.

Liberman, A., and Mattingly, I. (**1985**). ''The motor theory of speech perception revisited,'' Cognition **21**, 1–36.

Lyxell, B., Andersson, J., Andersson, U., Arlinger, S., Bredberg, G., and Harder, H. (**1998**). ''Phonological representation and speech understanding with cochlear implants in deafened adults,'' Scand. J. Psychol. **39**, 175–179.

McKay, C. M., and Henshall, K. R. (**2002**). ''Frequency-to-electrode allocation and speech perception with cochlear implants,'' J. Acoust. Soc. Am. **111**, 1036–1044.

McKay, C. M., and McDermott, H. J. (**1993**). ''Perceptual performance of subjects with cochlear implants using the Spectral Maxima Sound Processor (SMSP) and the Mini Speech Processor (MSP),'' Ear Hear. **14**, 350–367.

Osberger, M. J., and Fisher, L. (**1999**). ''SAS–CIS preference study in postlingually deafened adults implanted with the Clarion cochlear implant,'' Ann. Otol. Rhinol. Laryngol. **177** (Suppl), 74–79.

Sarant, J. Z., Blamey, P. J., Cowan, R. S., and Clark, G. M. (**1997**). ''The effect of language knowledge on speech perception: What are we really assessing?'' Am. J. Otol. **18**, (Suppl), 135–137.

Skinner, M., Fourakis, M., Holden, T., Holden, L., and Demorest, M. (**1996**). ''Identification of speech by cochlear implant recipients with the Multipeak (MPEAK) and Spectral Peak (SPEAK) speech coding strategies. I. Vowels,'' Ear Hear. **17**, 182–197.

Skinner, M., Fourakis, M., Holden, T., Holden, L., and Demorest, M. (**1999**). ''Identification of speech by cochlear implant recipients with the Multipeak (MPEAK) and Spectral Peak (SPEAK) speech coding strategies. II. Consonants,'' Ear Hear. **20**, 443–460.

Tyler, R. S., Gantz, B. J., Woodworth, G. G., Parkinson, A. J., Lowder, M. W., and Schum, L. K. (**1996**). ''Initial independent results with the Clarion cochlear implant,'' Ear Hear. **17**, 528–536.

Tyler, R. S., and Moore, B. C. (**1992**). ''Consonant recognition by some of the better cochlear-implant patients,'' J. Acoust. Soc. Am. **92**, 3068–3077.

Tyler, R. S., Preece, J. P., Lansing, C. R., and Gantz, B. J. (**1992**). ''Natural vowel perception by patients with the Ineraid cochlear implant,'' Audiology **31**, 228–239.

Van Tasell, D. J., Greenfield, D. G., Logemann, J. J., and Nelson, D. A. (**1992**). ''Temporal cues for consonant recognition: Training, talker generalization, and use in evaluation of cochlear implants,'' J. Acoust. Soc. Am. **92**, 1247–1257.

Van Wiringen, A., and Wouters, J. (**1999**). ''Natural vowel and consonant recognition by Laura cochlear implantees,'' Ear Hear. **20**, 89–103.

Vandali, A. (**2001**). ''Emphasis of short-duration acoustic speech cues for cochlear implant users,'' J. Acoust. Soc. Am. **109**, 2049–2061.

Wang, M., and Bilger, R. (**1973**). ''Consonant confusions in noise: A study of perceptual features,'' J. Acoust. Soc. Am. **54**, 1248–1266.

Wilson, B. (**2000**). *Sixth Quarterly Progress Report, Speech Processors for Auditory Prostheses* (Research Triangle Institute, Research Triangle Park, NC).