

Evaluating the function of phonetic perceptual phenomena within speech recognition: An examination of the perception of /d/-/t/ by adult cochlear implant users

Paul Iverson^{a)}

Department of Phonetics and Linguistics, University College London, London NW1 2HE, England and Department of Otolaryngology—Head and Neck Surgery, University of Iowa Hospitals and Clinics, Iowa City, Iowa 52242-1078

(Received 5 July 2002; revised 28 September 2002; accepted 4 November 2002)

This study examined whether cochlear implant users must perceive differences along phonetic continua in the same way as do normal hearing listeners (i.e., sharp identification functions, poor within-category sensitivity, high between-category sensitivity) in order to recognize speech accurately. Adult postlingually deafened cochlear implant users, who were heterogeneous in terms of their implants and processing strategies, were tested on two phonetic perception tasks using a synthetic /da/-/ta/ continuum (phoneme identification and discrimination) and two speech recognition tasks using natural recordings from ten talkers (open-set word recognition and forced-choice /d/-/t/ recognition). Cochlear implant users tended to have identification boundaries and sensitivity peaks at voice onset times (VOT) that were longer than found for normal-hearing individuals. Sensitivity peak locations were significantly correlated with individual differences in cochlear implant performance; individuals who had a /d/-/t/ sensitivity peak near normal-hearing peak locations were most accurate at recognizing natural recordings of words and syllables. However, speech recognition was not strongly related to identification boundary locations or to overall levels of discrimination performance. The results suggest that perceptual sensitivity affects speech recognition accuracy, but that many cochlear implant users are able to accurately recognize speech without having typical normal-hearing patterns of phonetic perception. © 2003 Acoustical Society of America. [DOI: 10.1121/1.1531985]

PACS numbers: 43.71.Es, 43.71.Ky [KG]

I. INTRODUCTION

The ability of individuals to recognize speech via cochlear implants calls for a reconsideration of what types of phonetic information and perceptual processing are necessary for human speech recognition. Cochlear implants bypass much of the auditory periphery, such that the neural firing patterns resulting from cochlear implant stimulation differ from normal neural firing patterns (e.g., Rubenstein *et al.*, 1999). The functional number of frequency channels is fewer than for normal hearing (e.g., Dorman *et al.*, 2000; Fishman *et al.*, 1997), but temporal resolution can be about the same (e.g., Busby *et al.*, 1993; Shannon, 1989, 1992; see Shannon, 1993 for a review). Despite the facts that cochlear implant stimulation is quite different from normal hearing in many respects, and that the standard frequency-related phonetic cues (e.g., formant and burst frequencies) may be difficult to discern given the poor spectral resolution of cochlear implants (e.g., Dorman, 1991; Shannon *et al.*, 1995), the best postlingually deafened users of current cochlear implants are able to recognize more than 90% words correct in clinical tests of open-set sentence recognition (e.g., Parkinson *et al.*, 1998).

The aim of the present study was to determine whether cochlear implant users must perceive differences along phonetic continua in the same way as do normal hearing listen-

ers (e.g., having sharp identification boundaries, low within-category sensitivity, and high between-category sensitivity; Liberman *et al.*, 1957; Studdert-Kennedy *et al.*, 1970; Repp, 1984) in order to recognize speech accurately. Dorman and colleagues (Dorman *et al.*, 1991) found that, among a group of six Symbion cochlear implant users who had above-average word recognition accuracy, four had phoneme labeling functions for a synthetic voice-onset-time (VOT) continuum that were like those of normal-hearing individuals. It is thus clear that at least a subset of cochlear implant users is similar to normal-hearing individuals, but it is unlikely that all cochlear implant users perceive phonetic differences in this way (cf. Hedrick and Carney, 1997), particularly given their large range of individual differences in speech recognition (e.g., Parkinson *et al.*, 1998). It is unknown whether the cochlear implant users with normal phoneme identification functions are more accurate at recognizing speech, or whether it is possible for individuals to recognize speech accurately despite having unusual patterns of phonetic perception.

In the normal-hearing speech recognition literature, current evidence contradicts the early interpretation of categorical perception, that speech is perceived in terms of phoneme labels (e.g., Liberman *et al.*, 1967). For example, fine-grained phonetic variation (e.g., variation due to differences between talkers) has been shown to affect speech recognition accuracy and to be stored in memory (e.g., Goldinger, 1996; Nygaard and Pisoni, 1998; Pisoni, 1997); listeners have been

^{a)}Electronic mail: paul@phon.ucl.ac.uk

shown to perceive differences in goodness among stimuli that are categorized the same (e.g., Allen and Miller, 2001; Iverson and Kuhl, 1995, 1996, 2000; Miller, 1994); and current cognitive models of word recognition have been able to account for experimental data without including a phoneme categorization stage (e.g., Connine *et al.*, 1994; Luce and Pisoni, 1998; Norris *et al.*, 2000). Despite the fact that phoneme encoding may not occur, the perceptual phenomena associated with the categorical perception of consonants (e.g., sensitivity peaks near identification boundaries) have remained robust and ubiquitous in the literature. There has been little direct evidence, however, to indicate what role these perceptual phenomena have in speech recognition.

It is difficult to address this issue by testing normal-hearing individuals listening to their native language, because few individuals have unusual patterns of phonetic perception but are normal in other respects. However, language experience can produce these types of individual differences, and cross-language studies have consistently linked individual differences in phonetic perception and word recognition. For example, Japanese adults who have difficulty identifying synthetic /r/-/l/ syllables also have difficulty recognizing words with those phonemes (Yamada, 1995), and non-native speakers of English have a marked difficulty recognizing English words that require more phonetic information to be distinguished from lexical competitors (Bradlow and Pisoni, 1999). One drawback of cross-language research is that the origin of these speech recognition difficulties cannot be definitively isolated to any one level, because language experience affects many levels of neural processing simultaneously. The present study examined postlingually deafened adults with cochlear implants, because their speech recognition difficulties have a clearer sensory origin, and it can be assumed that these individuals have normal native-language linguistic processing due to their hearing during childhood.

The experiments measured phonetic perception (identification and discrimination) along a /da/-/ta/ synthetic continuum, and speech recognition (open-set word recognition and forced-choice phoneme identification) for recordings of natural speech from multiple talkers. Phonetic perception experiments have traditionally used fixed-interval designs (e.g., 10-ms VOT differences between all pairs of stimuli in a discrimination task). Such designs are likely inappropriate for cochlear implant users given their wide range of individual differences (i.e., for the same interval sizes, better subjects would reach ceiling performance in discrimination tasks and poorer subjects would be at chance). Instead, the present experiments used adaptive procedures (Levitt, 1971) to adjust the interval size for individual subjects. Measures of phonetic perception along the /da/-/ta/ synthetic continuum were compared to those of normal-hearing individuals, to assess whether the normality of phonetic perception for these stimuli is predictive of individual differences in speech recognition performance.

II. METHOD

A. Subjects

Twenty-five postlingually deafened cochlear implant users were tested. The subjects were not selected based on implant type or processing strategy, to increase the potential individual differences among subjects; eight used the Clarion implant with a CIS processing strategy, one used an Ineraid implant with a Med-El processor and a CIS processing strategy, six used a Nucleus-22 implant with a SPEAK processing strategy, four used a Nucleus-24 implant with an ACE processing strategy, five used a Nucleus-24 implant with a SPEAK processing strategy, and one had binaural Nucleus-24 implants, one with SPEAK and the other with ACE. The age of the subjects had a range of 40.8–80.3 years, with a mean of 58.6 years. Their duration of implant use had a range of 0.5–12.2 years with a mean of 5.9 years. Fourteen cochlear implant subjects were male and 11 were female. All were native speakers of American English.

Fourteen normal-hearing subjects were tested to provide comparison data on the phonetic perception tasks. Two subjects were dropped from this study because of unusual data; one subject had no clear sensitivity peak in the discrimination task (discrimination was accurate in a broad region near the identification boundary), and the other had levels of discrimination performance that were more than 2 standard deviations poorer than the average. These unusual data were omitted because they were not consistent with the aim of estimating typical normal-hearing performance. The age of the 12 remaining normal-hearing subjects had a range of 21.1–56.0 years, with a mean of 33.3 years. Four of these subjects were male and eight were female. All were native speakers of American English.

B. Apparatus

The subjects were tested in a double-walled booth. The stimuli were presented at a comfortable level via a computer sound card connected to two loudspeakers, positioned to the front-left and front-right of the subjects. Subjects entered their responses by clicking on buttons displayed on a computer screen, using a computer mouse. One subject was blind, and used a modified testing interface that collected responses via a button box.

C. Stimuli

1. Natural recordings

A list of 120 monosyllabic words and 80 /da/ and /ta/ syllables was recorded by ten adult native speakers of American English who lived in Iowa. Five talkers were male and five were female. The word corpus comprised 20 /d/-/t/ minimal pairs (i.e., 40 words) with the target phonemes in syllable-initial position (*target-initial words*), 20 /d/-/t/ minimal pairs with the target phonemes in syllable-final position (*target-final words*), and 40 words that did not contain either /t/ or /d/ and were randomly selected from The Celex Lexical Database (1995; *nontarget words*). Minimal pairs were used for the target words so that the lexicon could not be used to distinguish /d/ and /t/ during the word recognition experi-

ment. The nontarget words were included in the corpus so that responses in the word recognition experiment would be less likely to be biased toward words containing /d/ or /t/.

During recording, the words and syllables were displayed one at a time on a computer screen, in a random order. The words were recorded using 16-bit samples and a 44.1-kHz sampling rate.

The recordings were screened for intelligibility and recording quality. The amplitude of each recording was scaled to make all recordings equal in rms amplitude. The final word corpus was selected to include 4 target-initial words (2 /d/ and 2 /t/), 4 target-final words (2 /d/ and 2 /t/), and 4 nontarget words from each of the ten talkers. Each of the 120 words occurred once in the final corpus. The final syllable corpus included 4 /da/ and 4 /ta/ syllables from each of the ten talkers.

VOT was measured for the initial-target phonemes, quantified here as the latency between burst onset and the onset of voicing energy in the $F2$ range (i.e., onset of regular voicing). In words, the /d/ phonemes had an average VOT of 27 ms and a range of 10–51 ms, excluding two prevoiced stimuli; the /t/ phonemes had an average VOT of 104 ms and a range of 56–148 ms. In syllables, the /d/ phonemes had an average VOT of 23 ms and a range of 13–40 ms, excluding one prevoiced stimulus; the /t/ phonemes had an average VOT of 94 ms and a range of 48–136 ms.

2. Synthetic continuum

The stimulus continuum was created using the Klatt synthesizer controlled by higher-level articulatory parameters within the HLSYN computer program (1997; Stevens and Bickley, 1991). The synthesis parameters (e.g., formant frequencies and fundamental frequency contour) were modeled from recordings of /da/ and /ta/ by a male speaker. The duration of voicing was 350 ms for every stimulus (i.e., the aspirated portion of each stimulus was added to the total stimulus length, rather than subtracted from the duration of the voiced portion). The formant frequencies for $F1$ – $F4$ at the consonant release were 200, 1762, 2889, and 2972 Hz. The frequencies of $F1$ – $F4$ at the vowel target were 781, 1501, 2532, and 3029 Hz. $F0$ fell from 120 to 80 Hz during the voiced portion of the stimuli. An articulatory parameter representing the cross-sectional area of a constriction formed at the tongue blade (ab) was set to 0 mm² during the consonant closure and reached 100 mm² (i.e., no constriction) 10 ms after the release of the closure.

VOT ranged from 0–150 ms, with a step size of 1 ms (i.e., a total of 151 stimuli). This variation in VOT was created by manipulating an articulatory parameter for the area of glottal opening (ag), relative to the release of the consonant closure (i.e., the start of the transition of ab from 0 to 100 mm²). For example, a stimulus with a 0-ms VOT had modal voicing ($ag = 5$ mm²) beginning at the same time as the closure release. A stimulus with a 100-ms VOT had aspiration at the closure release ($ag = 30$ mm²) and modal voicing ($ag = 5$ mm²) 100 ms after the closure release. As a consequence of manipulating these articulatory parameters (i.e., ag and ab), multiple acoustic cues, such as the latency between the burst and voicing, the burst amplitude, and the

$F1$ onset, were all varied according to Hlsyn's (1997) articulatory model. The acoustic cues for VOT thus were designed to vary naturally along the stimulus continuum, and they were not directly controlled to equate acoustic differences among stimuli.

D. Procedure

The four experimental tasks were run in a single session for each subject, in the order listed below. Subjects were allowed to take breaks between experimental tasks.

1. Open-set word recognition

Subjects heard one word on each trial and identified what they thought they heard. Subjects were given the option to either type their response into the computer or tell the experimenter the word that they heard. Subjects were instructed that all of the stimuli would be real monosyllabic words, and that they needed to type their best guess for the word even if they were not certain. Subjects were not told that the word corpus had a high percentage of words containing /t/ or /d/. Moreover, this was the first condition that was run for each subject, and subjects had yet to be told that the later conditions would involve /t/ and /d/ identification. Postexperiment comments by the subjects suggested that they were unaware that there were a large number of /t/ and /d/ words in the corpus, although some subjects noticed that some of the words rhymed.

Each of the 120 words was presented in an order that was randomized for each subject. There was no practice or feedback.

After the experiment was completed, each response was corrected for spelling and transcribed phonemically. The responses were scored in terms of whether the word response was correct and whether the target phoneme was correct.

2. Phoneme identification: Natural syllables

Subjects heard one syllable on each trial and judged whether it began with /d/ or /t/. Subjects began with a short practice session composed of randomly selected trials with no feedback; the practice was ended as soon as the experimenter and the subject were confident that the subject was comfortable with the task. Subjects then completed an experimental session composed of the full corpus of 80 syllables presented in a random order for each subject.

3. Phoneme identification: Synthetic continuum

As with natural syllables, subjects were presented one stimulus on each trial and judged whether it began with /d/ or /t/. Subjects began with a short practice session composed of randomly selected trials (from 0 to 120-ms VOT) with no feedback. The practice was ended as soon as the experimenter and the subject were confident that the subject was comfortable with the task.

In the experimental session, the trials were set by an interleaved double-staircase adaptive procedure that was designed to find the identification boundary location and width for each subject. Specifically, one-up/two-down Levitt procedures (1971) were used to find two locations along the

stimulus continuum: The point where stimuli were identified as /d/ on 71% of trials (found by the /d/ series of the adaptive procedure), and the point where stimuli were identified as /t/ on 71% of trials (found by the /t/ series of the adaptive procedure). The midpoint between these locations was defined as the identification boundary location. The difference between these locations was defined as the identification boundary width.

The adaptive procedure had four stages. In the first stage, the /d/ series began with a 16-ms VOT and the /t/ series began with a 54-ms VOT. The step size was 16 ms, and the first stage was completed after both adaptive series completed three reversals. The second stage had a step size of 8 ms and was finished when both series completed seven reversals. The third stage began by resetting the values of /d/ and /t/ to the average of their reversals in the second stage, had a step size equal to half the difference between the /d/ and /t/ series (estimated from stage 2), and was finished when both series completed 11 reversals. The fourth stage began by resetting the values of /d/ and /t/ to the average of their reversals in the third stage, had a step size equal to half the difference between the /d/ and /t/ series (estimated from stage 3), and was finished after both series completed 15 reversals. The average of the reversals in stages 2–4 were used to calculate the boundary locations.

Half of the presented trials were from neither adaptive series. On these trials subjects were presented a stimulus that was randomly selected from the series, to prevent the responses from being affected in some way by having stimuli concentrated only at the phoneme boundary. The results from these trials were not included in the estimation of boundary locations.

The order of all trials (i.e., /d/ series, /t/ series, and the other trials) was randomized for each subject.

4. Discrimination

Subjects heard three stimuli on each trial with an inter-stimulus interval of 250 ms. Two stimuli were the same and one was different, and the different stimulus was either the first or last that they heard. Subjects gave a two-alternative forced-choice response to indicate which stimulus, the first or the last, they thought was different.

Discrimination was tested at 14 different anchor points along the synthetic stimulus continuum: The locations of the identification boundary, the best¹ /d/, and the best /t/ for each subject; and at the points 0, 10, 20, 30, 40, 50, 60, 70, 80, 100, and 120-ms VOT. A one-up/two-down Levitt procedure (1971) was used for each anchor point to find the amount of VOT difference between stimuli that was required to perform the task at 71% correct (the 14 adaptive series were run within the same blocks of trials).

The stimuli were centered around each anchor point. For example, if the anchor point was 50 ms and the difference between stimuli was 10 ms, then subjects were tested with 45- and 55-ms stimuli. The stimulus selection was altered when an end of the range (0 or 150 ms) was reached. For example, if an anchor point was 10 ms and the VOT difference was 30 ms, subjects were tested with 0- and 30-ms stimuli.

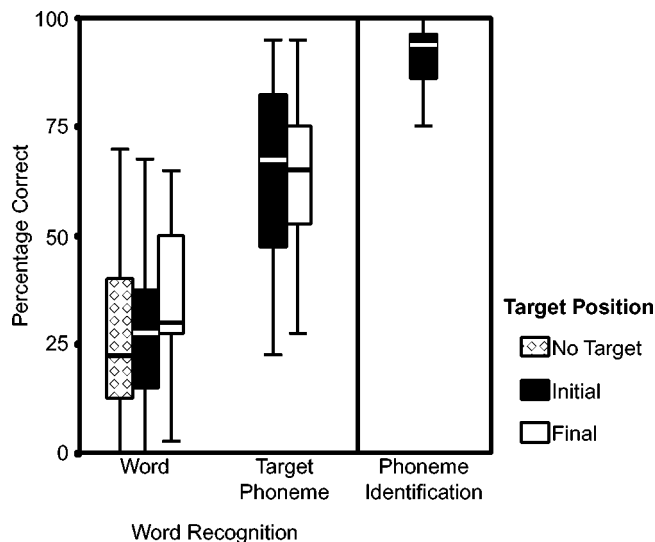


FIG. 1. Boxplots of results for the open-set word recognition and phoneme identification tasks using natural recordings of speech. Boxplots display the interquartile range of scores. The box shows the 25th to 75th percentiles, with a line at the median value. The lower and upper “whiskers,” respectively, show the first and last quartiles. Nontarget words were those that did not have either /t/ or /d/. Target-initial words had either /t/ or /d/ in syllable initial position. Target-final words had either /t/ or /d/ in syllable final position.

The adaptive procedure varied VOT multiplicatively. For example, if the step size was 2, the VOT difference between stimuli was doubled after an incorrect response and halved after two correct responses. The VOT difference was limited so that it was never less than 1 ms or greater than 100 ms.

Subjects completed a practice, without feedback, in which they heard a randomly selected anchor point and a randomly selected VOT difference on each trial. The practice was terminated as soon as the subject and the experimenter were confident that the subject understood the task.

The experimental session had seven stages. In stage 1, the VOT differences began at 16 ms, the adaptive step size was 2, and there were two reversals for each anchor point. In stages 2–4, the VOT difference for each anchor point was reset at the beginning of the stage to be equal to the average reversals at that anchor point and at the adjacent anchor points along the series (the adjacent anchor points entered into this calculation because they provided additional information about sensitivity at that general location in the VOT continuum). The adaptive step size was $2^{0.5}$ and there were two reversals at each stage. Stages 5–7 had an adaptive step size of $2^{0.25}$, but were the same as stages 2–4 in all other respects. Subjects were permitted to take a short break between stages.

The difference limen (DL) at each anchor point for each subject was calculated by averaging the two median reversals in stages 2–7. The location of the sensitivity peak (i.e., minimum DL) for each subject was estimated using parabolic interpolation (Press *et al.*, 1992). Specifically, the sensitivity peak location was defined to be the minimum of a parabola that was found by the equation

$$\min = \frac{b - 0.5 * \{ [b - a]^2 * [f(b) - f(c)] - [b - c]^2 * [f(b) - f(a)] \}}{[b - a] * [f(b) - f(c)] - [b - c] * [f(b) - f(a)]}, \quad (1)$$

where b was the anchor point with the lowest measured DL, a and c were the anchor points adjacent to b , and $f(a)$, $f(b)$, and $f(c)$ were the DL values at these anchor points. Interpolation was used so that sensitivity peaks were based on the data from three points rather than one, and so that the location estimates had a higher resolution than did the anchor point locations.

III. RESULTS

A. Word recognition and phoneme identification for natural recordings

Figure 1 displays the ranges of word recognition and phoneme identification results for cochlear implant subjects. As is typical of cochlear implant users, there was substantial individual variability in percentage-correct scores for entire words and for target phonemes within words. Their average word recognition accuracy was poor (average of 29%), but this likely was due to the difficulty of this particular word corpus; these subjects had averaged 59% correct for CNC words, in tests conducted during their clinical visits. The percentage-correct scores in the forced-choice syllable identification task approached ceiling levels of performance (i.e., 100%), which reduced the range of scores.

B. Phoneme identification and discrimination: Synthetic stimulus continuum

1. Sensitivity functions

Figure 2 displays *sensitivity functions* (i.e., DL values along the VOT continuum) and identification boundary loca-

tions. The normal-hearing subjects were relatively homogeneous. Sensitivity was best (i.e., lowest DL) in the region of 30–40 ms along the stimulus series, near the average normal-hearing category boundary (37 ms). Sensitivity was poorest within phoneme categories.

The sensitivity functions from cochlear implant subjects were highly variable, to an extent that would make the presentation of group sensitivity functions meaningless. Instead, sensitivity functions from three individual subjects are displayed in Fig. 2. Subject 1 is an example of a cochlear implant user who had results that were similar to those of normal-hearing subjects. There was a clear sensitivity peak near the category boundary (at a longer VOT than was found for normal-hearing subjects), and poorer sensitivity within phoneme categories. At the sensitivity peak, the level of sensitivity was within the normal-hearing range. Within phoneme categories, sensitivity was somewhat poorer than was found for normal-hearing subjects.

However, many subjects had data that were markedly different from that of normal-hearing individuals. For example, Subject 2 did not have an identification boundary and a sensitivity peak at the same location. This subject had an identification boundary that was near that found for normal-hearing individuals, but had a sensitivity peak at a lower VOT (14 ms) and perhaps a second sensitivity peak at a higher VOT (80 ms). The subject also had much poorer sensitivity overall compared to normal-hearing subjects, and approached the 100-ms maximum difference at several points along the continuum. This makes the sensitivity peak difficult to interpret, because large DLs should lead to more

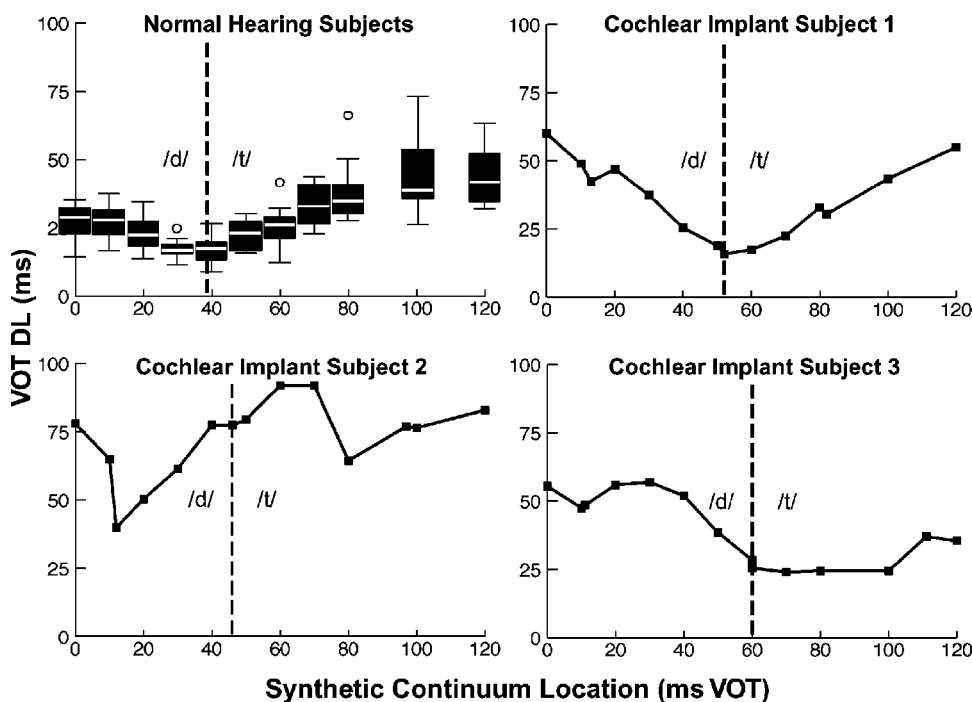


FIG. 2. Boxplots of sensitivity functions for normal-hearing subjects and individual sensitivity functions for three example cochlear implant subjects. Boxplots display the interquartile range of scores, with outliers marked with circles. The vertical dashed lines in each plot indicate the location of the phoneme identification boundary. The normal-hearing subjects were fairly homogeneous and had results consistent with categorical perception (i.e., high sensitivity at the category boundary, low sensitivity within phoneme categories). The data from the cochlear implant subjects were highly variable; there is evidence of categorical perception for Subject 1, but not for Subjects 2 and 3.

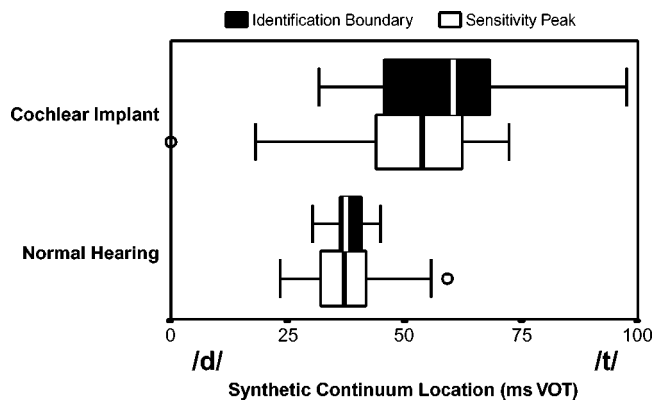


FIG. 3. Boxplots of the locations of identification boundaries and sensitivity peaks along the synthesized stimulus continuum. The distributions of both measures were shifted to longer VOT values for cochlear implant users, compared to those of normal-hearing individuals. Moreover, the individual differences were greater for cochlear implant users than for normal-hearing individuals, on both location measures.

gradual changes along the continuum (because neighboring stimulus pairs have more overlap); this individual had very sharp changes in the sensitivity function near the peak.

Subject 3 is an example of an intermediate case. Sensitivity was poor within the /d/ category and increased near the category boundary, but sensitivity within the /t/ category remained as high as at the category boundary, forming a broad region of high sensitivity rather than a peak. In fact, this individual had higher sensitivity for stimuli within the /t/ category (i.e., 60–120 ms on the continuum) than did any of the normal-hearing subjects.

2. Location measures: Sensitivity peaks and identification boundaries

As displayed in Fig. 3, cochlear implant subjects tended to have sensitivity peaks and identification boundaries at longer VOT values than did normal-hearing subjects. Cochlear implant subjects also had a wider range of VOT locations for both measures, such that some cochlear implant users had identification boundary and sensitivity peak locations that were within, or below, the normal-hearing range.

The correlation between the locations of sensitivity peaks and category boundaries for cochlear implant users was significant, $r=0.49$, $p<0.01$. However, few cochlear implant users had their sensitivity peak and identification boundary at exactly the same location; the difference between the two location measures was as large as 49.9 ms for one subject, and there was a median difference among subjects of 15.3 ms. There appeared to be continuous variation among subjects in the extent to which the locations of sensitivity peaks and identification boundaries differed.

3. Sensitivity measures: Minimum DLs and identification boundary widths

As displayed in Fig. 4, cochlear implant subjects had larger identification boundary widths and larger minimum DL values than did normal-hearing subjects. Although there was some overlap between these distributions, it appears that, as a group, cochlear implant users are less sensitive, compared to normal-hearing individuals, to changes in VOT

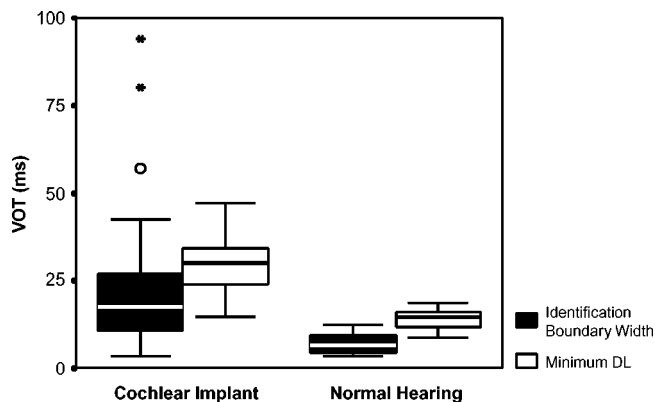


FIG. 4. Boxplots of identification boundary widths and minimum DL values. Although there is overlap between the distributions for individuals with cochlear implants and normal hearing, the cochlear implant users generally had greater widths and DL minima, demonstrating that they were less sensitive to VOT differences near their /d-/t/ phoneme boundary.

near identification boundaries and sensitivity peaks. These two measures were correlated for cochlear implant users, $r=0.47$, $p<0.01$.

C. Relationships among experimental measures

Initial analysis of the data suggested an inverted U-shaped relationship between the location measures (identification boundary and sensitivity peak) and speech recognition measures for cochlear implant users. In Fig. 5, for example, there is a significant curvilinear relationship, measured using polynomial regression, between sensitivity peak location and initial target phoneme recognition within words, $R=0.65$, $F(21)=5.16$, $p<0.01$. This shows that subjects who had sensitivity peaks near 45–50 ms along the stimulus series had the highest phoneme recognition accuracy, and accuracy declined for subjects who had sensitivity peaks at longer or shorter VOT values. To allow for simpler linear statistical comparisons with the speech recognition scores, the location measures were recalculated in terms of their distance from the peak location of the inverted U-shaped function (i.e., 47.5 ms). These recalculated mea-

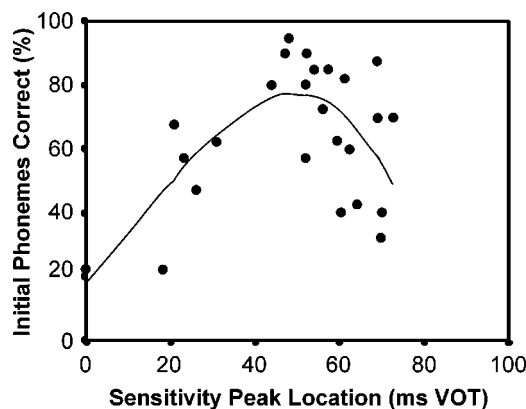


FIG. 5. Scatterplot of the relationship between sensitivity peak locations and the percentage-correct initial target phoneme recognition within words. There was a significant inverted-U-shaped relationship, indicated by the best-fit polynomial regression line, between sensitivity peak location and phoneme recognition; subjects with a sensitivity peak at a 45–50-ms VOT tended to have higher phoneme recognition scores.

TABLE I. Correlations (r) of measures of word recognition and phonetic perception for cochlear implant subjects.

	Words			Phonemes in words		Forced-choice identification
	Nontarget	Initial	Final	Initial	Final	
Optimality of identification boundary location	-0.27	-0.26	-0.21	-0.31	-0.09	-0.24
Optimality of sensitivity peak location	-0.38 ^a	-0.47 ^a	-0.47 ^a	-0.70 ^a	-0.45 ^a	-0.53 ^a
Identification boundary width	-0.32	-0.29	-0.45 ^a	-0.28	-0.37 ^a	-0.16
Minimum DL	-0.29	-0.16	-0.10	-0.06	-0.18	-0.11

^a $p < 0.05$.

asures thus quantified how far each listeners' identification boundary and sensitivity peak locations were from the optimal location for word recognition.²

Pearson correlation coefficients between these phonetic perception and speech recognition measures are displayed in Table I. The results demonstrated that there was a clear and consistent relationship between the optimality of sensitivity peak locations and all speech recognition measures. There was a particularly strong tendency ($r = -0.70$) for individuals with sensitivity peaks near 47.5 ms to correctly recognize initial target phonemes within words, and there was even a tendency ($r = -0.38$) for these individuals to correctly recognize words that did not contain /t/ or /d/. The relationship between identification boundary optimality and the speech recognition measures was weaker; although all correlations were in the expected negative direction, none was significant.

The correlations (Table I) revealed that there was a weak inverse relationship between identification boundary width and the speech recognition measures, reaching significance only for target-final words and phonemes; individuals with a broader phoneme identification boundary tended to have more difficulty recognizing these phonemes within natural speech. In contrast, minimum DL was not significantly correlated with any of the speech recognition measures. It is somewhat surprising that identification boundary width was more strongly related to word recognition than was minimum DL, because both are sensitivity measures (i.e., sharper identification boundaries indicate higher sensitivity, as do smaller DLs). However, the identification boundary width can also be interpreted as an indirect measure of the optimality of the identification boundary location, because identification boundary widths can be expected to be sharper when the identification boundaries and sensitivity peaks are at the same location. The minimum DL is more strongly related to the overall sensitivity to acoustic differences along the series.

To further test the contribution of all of the phonetic measures to word recognition accuracy, an ANCOVA analysis was conducted with non-, initial-, and final-target words coded as a repeated measure. Sensitivity peak location optimality was significant, $F(1,20) = 5.348$, $p < 0.05$. Identification boundary optimality, $F(1,20) = 1.039$, identification boundary width, $F(1,20) = 1.535$, and minimum DL, $F(1,20) = 0.078$, were not significant. Likewise, an ANCOVA was conducted for the phoneme recognition measures, with syllable identification, initial-target, and final-

target coded as a repeated measure. Again, sensitivity peak location optimality was significant, $F(1,20) = 13.112$, $p < 0.01$. Identification boundary optimality, $F(1,20) = 0.671$, identification boundary width, $F(1,20) = 1.756$, and minimum DL, $F(1,20) = 0.017$, were not significant. Together, these analyses confirm that sensitivity peak location optimality was the best predictor of speech recognition accuracy in this study.

The relationships between phonetic perception measures and speech recognition can be further illustrated by inspecting the example data presented in Fig. 3. Among these three subjects, word recognition performance was related to the shape of their sensitivity function; Subject 1 had high word recognition performance along with a normally shaped sensitivity function (e.g., 67.5% correct initial target words), and Subjects 2 and 3 had poorer word recognition performance (e.g., 30.0 and 32.5% correct initial-target words, respectively). These examples also illustrate why overall levels of sensitivity did not correlate with word recognition performance. Subject 3 had sensitivity levels that surpassed those of normal-hearing individuals within the /t/ category, but this did not lead to exceptional word recognition accuracy. Moreover, Subject 2 had levels of word recognition accuracy that were similar to those of Subject 3, despite Subject 2's much poorer levels of sensitivity.

IV. DISCUSSION

There were two main findings. First, cochlear implant users do not, as a group, perceive phonetic differences along a VOT continuum in the same way as do normal-hearing individuals; cochlear implant subjects tend to have sensitivity peaks and identification boundaries at longer VOT locations, identification boundaries that are less sharp, higher minimum DLs, and more intersubject variability on all of these measures. Second, speech recognition accuracy by cochlear implant users is related to the shape of the phonetic sensitivity function, at least in terms of the location of the peak, but is not strongly related to other aspects of phonetic perception, such as the level of sensitivity at the peak or to the phoneme identification boundary.

From the standpoint of normal-hearing speech perception theories, it is particularly notable that many cochlear implant subjects were able to accurately categorize voicing in natural speech (e.g., median forced-choice /d/-/t/ identification was 94%), despite the fact that their phonetic identi-

fication and sensitivity functions were markedly different from those of normal-hearing individuals. It is clearly not necessary to have normal categorical perception in order to recognize speech accurately. However, it is beneficial to have favorable sensitivity functions. When a listener has a sensitivity peak at an advantageous location (e.g., near normal-hearing identification boundaries), words likely become more distinct perceptually from potential lexical competitors, thereby facilitating recognition. Other types of sensitivity functions likely impair performance by making the perceptual differences between lexical competitors less salient than within-category variation. The patterns of sensitivity measured in phoneme discrimination experiments thus affect recognition accuracy, even though phoneme labeling may have little functional importance.

It was surprising that word recognition accuracy was unrelated to the overall level of sensitivity. Previous cochlear implant research has focused on the levels of spectral (e.g., Dorman *et al.*, 1996) or temporal (e.g., Cazals *et al.*, 1994; Hochmair-Desoyer *et al.*, 1985) resolution available to users (see also Svirsky, 2000). The present results provide a conflicting view; it seems more important for cochlear implant users to have relatively high sensitivity to critical VOT differences than it is for listeners to have high sensitivity to VOT differences throughout the continuum.

This conclusion is limited by the fact that it is based only on VOT data. It is logically necessary that word recognition performance must be affected by sensitivity levels to some extent, because accurate auditory word recognition would be impossible for individuals who were unable to hear any differences between sounds. Cochlear implant users, as a group, may have sensitivity levels for VOT that are above this lower limit, such that increases in phonetic sensitivity do not further improve recognition performance for voicing contrasts (see also Tyler *et al.*, 1989). The levels of sensitivity to VOT could prove important under more difficult listening conditions, such as when speech is combined with noise. Furthermore, the effects of sensitivity level could be stronger for phonetic dimensions that are more dependent on frequency cues (e.g., consonant place or vowel height), given that spectral sensitivity by cochlear implant users is generally poor.

It is unknown what caused the observed shifts in sensitivity peak locations. It would be straightforward to hypothesize that shifts of sensitivity peaks to longer VOTs are a result of temporal processing deficits. That is, normal-hearing research has suggested that VOT boundary locations could be a result of an auditory threshold for detecting the temporal order of a burst and the onset of voicing (e.g., Pastore and Farrington, 1996), so it would be reasonable to predict that individuals with poorer temporal resolution would have a higher threshold for detecting this difference, causing sensitivity peaks to occur at longer VOTs. However, there was no evidence that individuals with sensitivity peak locations at longer VOTs had unusually poor levels of sensitivity. Furthermore, this explanation does not account for why some individuals have sensitivity peaks at shorter-than-normal VOTs.

Sinex and colleagues (Sinex, McDonald, and Mott,

1991; cf. Soli, 1983) have suggested that spectral cues, such as the first formant frequency ($F1$), are more responsible for sensitivity peaks along VOT continua than are temporal cues. The locations of the sensitivity peaks along the continuum could thus have been affected by individual differences in $F1$ perception. For example, listeners may have been more sensitive to $F1$ transition differences when it spanned more than one electrode frequency band, or when it exceeded the low-frequency cutoff of the implant processor. The shifts in sensitivity peak locations along the VOT continuum could therefore have been caused by complex interactions between the characteristics of the cochlear implant processors, electrode locations, and the frequencies of the $F1$ transitions.

It is plausible too that cochlear implant users differ in their use of acoustic cues. Variability in cue weightings has been shown to occur among normal-hearing individuals (Hazan and Rosen, 1991), and the functional importance of these differences may be magnified when the available phonetic information is reduced. For example, individuals who attend to spectral cues for VOT (e.g., $F1$ onset) may have more difficulty discerning voicing through their cochlear implant than do individuals who attend to temporal cues for VOT (e.g., duration of aspiration), which tend to be better represented via cochlear implants. Individual differences in cue weightings may be particularly large for cochlear implant users, arising from changes to speech recognition strategies following prolonged periods of deafness and a subsequent accommodation to electric hearing.

ACKNOWLEDGMENTS

This work was supported by research Grants Nos. 1 R03 DC03999 and 2 P50 CD 00242 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health; and Grant No. RR00059 from the General Clinical Research Centers Program, Division of Research Resources, National Institutes of Health. I am grateful to Gina Hart and Annie Vranesic for their assistance with data collection; to Lynne E. Bernstein for initial comments on the research plan; and to Richard S. Tyler, Andrew Faulkner, and Stuart Rosen for their comments on this manuscript.

¹In a separate experiment, cochlear implant users were asked to rate the subjective goodness of the synthetic stimuli (see Iverson and Kuhl, 1995, 1996), and the stimuli with the highest goodness ratings for each category were used as anchor points in the discrimination task. The results from the goodness rating task are not discussed further, because of concerns over their reliability. Subjects mostly reported following the test that the stimuli “all sounded the same” or that they performed the goodness task on the basis of some idiosyncratic perceptual detail of the stimuli, such as loudness or “number of overtones.”

²Although the optimal location was operationally defined here as 47.5 ms, the average normal-hearing sensitivity peak location—which would have been expected to be optimal—occurred at a shorter VOT (37.7 ms). This discrepancy between potentially optimal locations could be due to not having enough statistical power to determine the exact shape of the relationship between location and recognition accuracy (e.g., there were no subjects who had sensitivity peaks between 33 and 43 ms). A function with a 37.7-ms optimal location could have fit the data nearly as well, if the slope of the function had been fit to be steeper toward the left than to the right.

Allen, J. S., and Miller, J. L. (2001). “Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate,” *Percept. Psychophys.* **63**, 798–810.

- Bradlow, A. R., and Pisoni, D. B. (1999). "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors," *J. Acoust. Soc. Am.* **106**, 2074–2085.
- Busby, P. A., Tong, Y. C., and Clark, G. M. (1993). "The perception of temporal modulations by cochlear implant patients," *J. Acoust. Soc. Am.* **94**, 124–131.
- Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C. (1994). "Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants," *J. Acoust. Soc. Am.* **96**, 2048–2054.
- Celex Lexical Database. (1995). (Version 2.5). Nijmegen: Center for Lexical Information, Max Planck Institute for Psycholinguistics.
- Connine, C. M., Blasko, D. G., and Wang, J. (1994). "Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context," *Percept. Psychophys.* **56**, 624–636.
- Dorman, M. F., Dankowski, K., McCandless, G., Parkin, J. L., and Smith, L. (1991). "Vowel and consonant recognition with the aid of a multichannel cochlear implant," *Q. J. Exp. Psychol. A* **43**, 585–601.
- Dorman, M. F., Loizou, P. C., Kemp, L. L., and Kirk, K. I. (2000). "Word recognition by children listening to speech processed into a small number of channels: Data from normal-hearing children and children with cochlear implants," *Ear Hear.* **21**, 590–596.
- Dorman, M. F., Smith, L. M., Smith, M., and Parkin, J. L. (1996). "Frequency discrimination and speech recognition by patients who use the Ineraid and continuous interleaved sampling cochlear-implant signal processors," *J. Acoust. Soc. Am.* **99**, 1174–1184.
- Fishman, K. E., Shannon, R. V., and Slattery, W. H. (1997). "Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor," *J. Speech Lang. Hear. Res.* **40**, 1201–1215.
- Goldinger, S. D. (1996). "Words and voices: Episodic traces in spoken word identification and recognition memory," *J. Exp. Psychol. Learn Mem. Cogn* **22**, 1166–1183.
- Hazan, V., and Rosen, S. (1991). "Individual variability in the perception of cues to place contrasts in initial stops," *Percept. Psychophys.* **49**, 187–200.
- Hedrick, M. S., and Carney, A. E. (1997). "Effect of relative amplitude and formant transitions on perception of place of articulation by adult listeners with cochlear implants," *J. Speech Lang. Hear. Res.* **40**, 1445–1457.
- HLSYN High-Level Parameter Speech Synthesis System. (1997). (Version 2.2). Sensimetics Corporation, Somerville, MA.
- Hochmair-Desoyer, I. J., Hochmair, E. S., and Stiglbrenner, H. K. (1985). "Psychoacoustic temporal processing and speech understanding in cochlear implant patients," in *Cochlear Implants*, edited by R. A. Schindler and M. M. Merzenich (Raven, New York), pp. 291–304.
- Iverson, P., and Kuhl, P. K. (1995). "Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling," *J. Acoust. Soc. Am.* **97**, 553–562.
- Iverson, P., and Kuhl, P. K. (1996). "Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/," *J. Acoust. Soc. Am.* **99**, 1130–1140.
- Iverson, P., and Kuhl, P. K. (2000). "Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism?" *Percept. Psychophys.* **62**, 874–886.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–471.
- Lieberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). "The discrimination of speech sounds within and across phoneme boundaries," *J. Exp. Psychol.* **54**, 358–368.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). "Perception of the speech code," *Psychol. Rev.* **74**, 431–461.
- Luce, P. A., and Pisoni, D. B. (1998). "Recognizing spoken words: The neighborhood activation model," *Ear Hear.* **19**, 1–36.
- Miller, J. L. (1994). "On the internal structure of phonetic categories: A progress report," *Cognition* **50**, 271–285.
- Norris, D. G., McQueen, J. M., and Cutler, A. (2000). "Merging information in speech recognition: Feedback is never necessary," *Behav. Brain Sci.* **23**, 299–325.
- Nygaard, L. C., and Pisoni, D. B. (1998). "Talker-specific learning in speech perception," *Percept. Psychophys.* **60**, 355–376.
- Parkinson, A. J., Parkinson, W. S., Tyler, R. S., Lowder, M. W., and Gantz, B. J. (1998). "Speech perception performance in experienced cochlear-implant patients receiving the SPEAK processing strategy in the Nucleus Spectra-22 cochlear implant," *J. Speech Lang. Hear. Res.* **41**, 1073–1087.
- Pastore, R. E., and Farrington, S. M. (1996). "Measuring the difference limen for identification of order of onset for complex auditory stimuli," *Percept. Psychophys.* **58**, 510–526.
- Pisoni, D. B. (1997). "Some thoughts on 'normalization' in speech perception," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego).
- Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. H. (1992). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge).
- Repp, B. (1984). "Categorical perception: Issues, methods, findings," in *Speech and Language*, edited by N. J. Lass (Academic, New York), Vol. 10, pp. 243–335.
- Rubinstein, J. T., Wilson, B. S., Finley, C. C., and Abbas, P. J. (1999). "Pseudospontaneous activity: Stochastic independence of auditory nerve fibers with electrical stimulation," *Hear. Res.* **127**, 108–118.
- Shannon, R. V. (1989). "Detection of gaps in sinusoids and pulse trains by patients with cochlear implants," *J. Acoust. Soc. Am.* **85**, 2587–2592.
- Shannon, R. V. (1992). "Temporal modulation transfer functions in patients with cochlear implants," *J. Acoust. Soc. Am.* **91**, 2156–2164.
- Shannon, R. V. (1993). "Psychophysics," in *Cochlear Implants: Audiological Foundations*, edited by R. S. Tyler (Singular, San Diego), pp. 357–388.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Sinex, D. G., McDonald, L. P., and Mott, J. B. (1991). "Neural correlates of nonmonotonic temporal acuity for voice onset time," *J. Acoust. Soc. Am.* **90**, 2441–2449.
- Soli, S. D. (1983). "The role of spectral cues in discrimination of voice onset time differences," *J. Acoust. Soc. Am.* **73**, 2150–2165.
- Stevens, K. N., and Bickley, C. A. (1991). "Constraints among parameters simplify control of Klatt formant synthesizer," *J. Phonetics* **19**, 161–174.
- Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., and Cooper, F. S. (1970). "Theoretical notes. Motor theory of speech perception: A reply to Lane's critical review," *Psychol. Rev.* **77**, 234–249.
- Svirsky, M. A. (2000). "Mathematical modeling of vowel perception by users of analog multichannel cochlear implants: Temporal and channel-amplitude cues," *J. Acoust. Soc. Am.* **107**, 1521–1529.
- Tyler, R. S., Moore, B. C., and Kuk, F. K. (1989). "Performance of some of the better cochlear-implant patients," *J. Speech Hear. Res.* **32**, 887–911.
- Yamada, R. A. (1995). "Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Timonium, MD), pp. 305–320.