Speech, Hearing and Language: work in progress

Volume 13

Individual differences in phonetic perception by adult cochlear implant users: effects of sensitivity to /d/-/t/ on word recognition

Paul IVERSON

Department of Phonetics and Linguistics
UNIVERSITY COLLEGE LONDON

Individual differences in phonetic perception by adult cochlear implant users: effects of sensitivity to /d/-/t/ on word recognition

Paul IVERSON

Abstract

This study examined whether the phonetic perceptual phenomena associated with categorical perception in normal-hearing listeners (i.e., sharp identification functions, poor within-category sensitivity, high between-category sensitivity) are predictive of individual differences in speech recognition performance among cochlear implant patients. Adult postlingually deafened cochlear implant users, who were heterogeneous in terms of their implants and processing strategies, were tested on 2 phonetic perception tasks using a synthetic /tɑ/-/dɑ/ continuum (phoneme identification and discrimination) and 2 speech recognition tasks using natural recordings from 10 talkers (open-set word recognition and forced-choice /t/-/d/ recognition). Cochlear implant users tended to have identification boundaries and sensitivity peaks at voice onset times (VOT) that were higher than found for normal-hearing individuals. Sensitivity peak locations correlated with individual differences in cochlear implant performance; individuals who had a /t/-/d/ 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 at least a subset of the perceptual phenomena associated with categorical perception have a functional role in word recognition by cochlear implant users.

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 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). This word recognition ability is particularly remarkable for postlingually deafened cochlear implant users. Adults are better at recognizing speech via cochlear implants if they had acoustic hearing during childhood (e.g., Naito et al., 1997; Tong et al., 1988; Waltzman et al., 1992), even though their electrical stimulation differs from the acoustic stimulation that guided their phonetic and lexical development.

The normal-hearing perceptual phenomena associated with the categorical perception of consonants (e.g., sharp identification boundaries, low within-category sensitivity, and high between-category sensitivity; Liberman et al., 1957; Studdert-Kennedy et al.,

1970; Repp, 1984) are foundations of speech science. But there is no *a priori* reason that cochlear implant users must exhibit these categorical perception phenomena in order to recognize words accurately. Even within the normal-hearing speech perception literature, there has been no clear evidence that categorical perception is functional or necessary for open-set word recognition. One could presume that categorical perception aids word recognition by eliminating superficial phonetic variability (via the assignment of phoneme labels), or by making stimulus words more distinct perceptually from lexical competitors. However, current cognitive models suggest that word recognition does not require discrete phoneme labeling (e.g., Connine et al., 1994; Luce and Pisoni, 1998; Norris et al., 2000). Furthermore, the ecological relevance of categorical perception data is questionable, because the synthesized continua that are typically used contain only a subset of the variability found in natural speech (see Johnson and Mullennix, 1997 for reviews) and forced-choice tasks may not engage the same cognitive processes involved in open-set word recognition (e.g., Sommers et al., 1997).

The present study examined phonetic perception and word recognition by postlingually deafened adult cochlear implant users. There were 4 experimental tasks: 2 phonetic perception tasks using a synthesized /dɑ/-/tɑ/ continuum (identification and discrimination), and 2 speech recognition tasks using recordings of natural speech (open-set word recognition and forced-choice phoneme identification).

The first goal of this study was to determine whether cochlear implant users exhibit the same types of identification and sensitivity functions along consonant continua that have been characteristic of speech perception by normal-hearing individuals[1]. Little has been published thus far on this issue. Dorman and colleagues have measured phoneme identification along /gɑ/-/kɑ/ continua by a single star user of the Symbion cochlear implant (Dorman et al., 1988), and by 6 users who had word recognition scores that were at least above median levels (Dorman et al., 1991). Of these 7 cochlear implant users, 5 had near-normal phoneme identification functions. It thus seems plausible that at least a subset of cochlear implant users exhibit categorical perception phenomena (c.f., Hedrick and Carney, 1997). However, it is unknown whether the identification results extend to individuals who use more recent multichannel implants, or have poorer word recognition abilities. Furthermore, discrimination along phonetic continua has not been tested.

The second goal was to see whether measures of phonetic perception can help account for individual differences in word recognition accuracy by cochlear implant users. Although the remarkable speech recognition abilities of the best cochlear implant users are as described above, it is also true that individual differences in performance are large. Some individuals obtain little phonetic information via their implant, scoring near 0% for words in sentences without lipreading (e.g., Parkinson et al., 1998). These individual differences, though currently not well understood, likely arise

1 The intent was not to literally test whether cochlear implant users have categorical perception, so there were no tests of whether phoneme identification predicted discrimination performance (in fact, the discrimination and identification data collected in this study precluded this type of analysis). Instead, the goal was to see more generally whether sensitivity along a phonetic continuum was similar in shape to that found for normal-hearing individuals (i.e., high-sensitivity near category boundaries and low-sensitivity within categories), and whether the phoneme identification boundaries were at similar locations to those found for normal-hearing individuals.

from multiple peripheral (e.g., electrical field interactions; Hanekom and Shannon, 1998), psychoacoustic (e.g., forward masking; Chatterjee, 1999), and cognitive (see Pisoni, 2000) sources. The present study examines whether measures of phonetic perception are subject to these large individual differences, and whether any particular phonetic measures are highly correlated with individual differences in speech recognition performance.

The third, and final, goal was to use this investigation of individual differences among cochlear implant users to provide clues about the basic perceptual and cognitive processes involved in speech recognition regardless of hearing status. To some extent, the uniformity of normal-hearing phonetic perception is a hindrance to understanding how people recognize speech. It is difficult to know, for example, whether having a phoneme identification boundary in the correct location along a phonetic continuum is important for word recognition, when all normal-hearing subjects with the same native language have nearly the same identification boundary location. Examining cochlear implant users can therefore be advantageous, because the large individual differences in speech recognition abilities may help reveal what aspects of phonetic perception have the most functional importance.

1. Method

1.1 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; 8 used the Clarion implant with a CIS processing strategy, 1 used an Ineraid implant with a Med-El processor and a CIS processing strategy, 6 used a Nucleus-22 implant with a SPEAK processing strategy, 4 used a Nucleus-24 implant with an ACE processing strategy, 5 used a Nucleus-24 implant with a SPEAK processing strategy, and 1 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 English speakers.

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, and the other had levels of discrimination performance that were more than 2 standard deviations poorer than the average. This unusual data was omitted because it was 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 8 were female. All were native English speakers.

1.2 Apparatus

The subjects were tested in a double-walled booth. The stimuli were produced by a computer sound card and were presented at a comfortable level via 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.

1.3 Stimuli

1.3.1 Natural Recordings

A list of 120 monosyllabic words and 80 /tɑ/ and /dɑ/ syllables were recorded by 10 adult native speakers of American English. Five talkers were male and 5 were female. The word corpus comprised 20 /t/-/d/ minimal pairs (i.e., 40 words) with the target phonemes in syllable-initial position (*target-initial words*), 20 /t/-/d/ 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; *non-target words*). Minimal pairs were used for the target words so that the lexicon could not be used to distinguish /t/ and /d/ during the word recognition experiment. The non-target words were included in the corpus so that responses in the word recognition experiment would be less likely to be biased toward words containing /t/ or /d/.

During recording, the words and syllables were displayed to the talkers individually on a computer screen (i.e., no sentence context), 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, and were equated in RMS amplitude. The final word corpus was selected to include 4 target-initial words (2 /t/ and 2 /d/), 4 target-final words (2 /t/ and 2 /d/), and 4 non-target words from each of the 10 talkers; each of the 120 words occurred once in the final corpus. The final syllable corpus included 4 /dɑ/ and 4 /tɑ/ syllables from each of the 10 talkers.

VOT was measured for the initial-target phonemes. In words, the /d/ phonemes had an average VOT of 27 ms and a range of 10-51 ms, excluding 2 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.

1.3.2 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). Spectrograms of example stimuli are displayed in Figure 1. The synthesis parameters (e.g., formant frequencies and fundamental frequency contour) were modeled from recordings of /dɑ/ and /tɑ/ 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 mm^2 during the consonant

closure and reached 100 mm^2 (i.e., no constriction) 10 ms after the release of the closure.

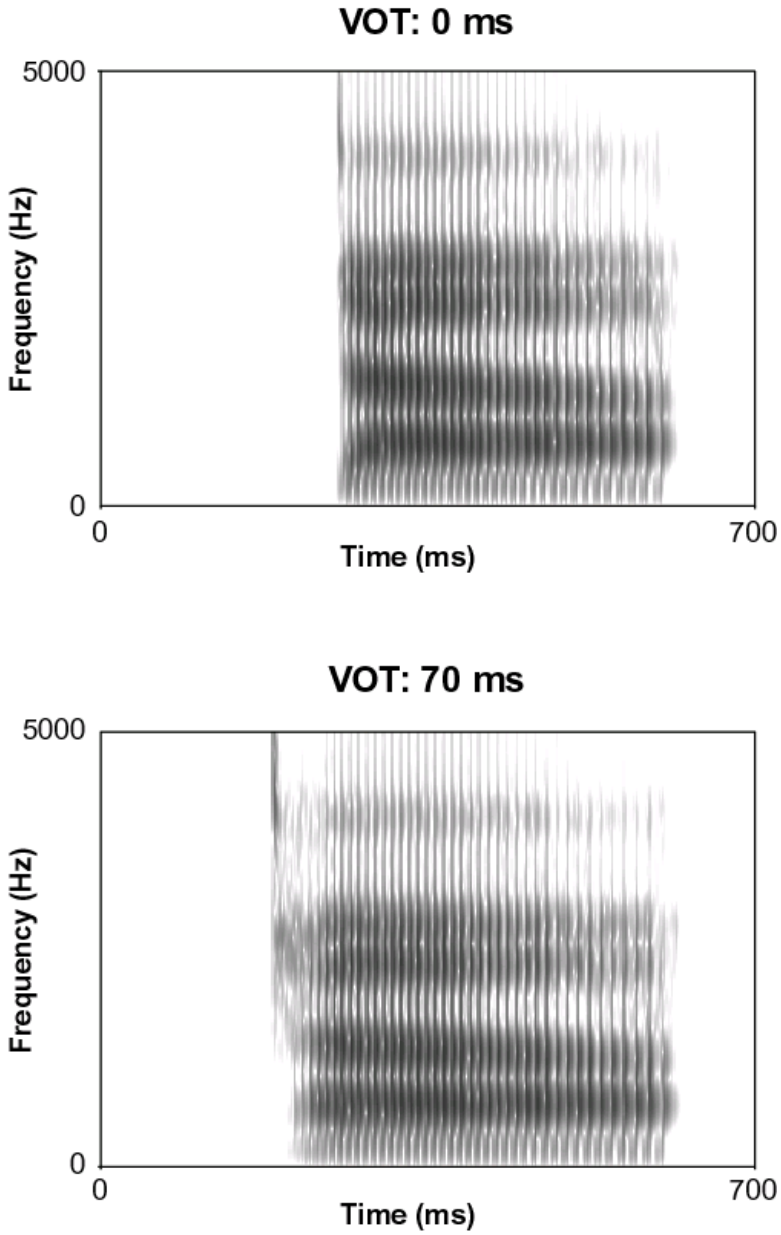

Figure 1: *Spectrograms of example synthetic stimuli. The 0-ms VOT stimulus was a clear example of /da/. The 70-ms VOT stimulus was a clear example of /ta/.*

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 mm^2). For example, a stimulus with a 0-ms VOT had modal voicing (ag = 5 mm^2) beginning at the same time as the closure release. A stimulus with a 100-ms VOT had aspiration at the closure release (ag = 30 mm^2) and modal voicing (ag = 5 mm^2) 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.

1.4 Procedure

The 4 experimental tasks were run in a single session (i.e., 2-3 hours) for each subject, in the order listed below. Subjects were allowed to take breaks between experimental tasks.

1.4.1 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. Post-experiment 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.

1.4.2 Phoneme Identification: Natural Syllables

Subjects heard one syllable on each trial and judged whether it began with /t/ or /d/. 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.

1.4.3 Phoneme Identification: Synthetic Continuum

As with natural syllables, subjects heard one stimulus on each trial and judged whether it began with /t/ or /d/. 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 4 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 3 reversals. The second stage had a step size of 8 ms and was finished when both series completed 7 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 heard 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.

1.4.4 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 best[2] /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.

2 In a separate experiment, cochlear implant users were asked to rate the subjective goodness of the synthetic stimuli (see Iverson and Kuhl, 1995, 1996). However, subjects were unable to reliably perform the goodness judgment task. Subjects mostly reported 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." There is no objective criteria for deciding which subjects could or could not perceive differences in goodness among these stimuli, so the results from this task were not considered further.

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 7 stages. In Stage 1, the VOT differences began at 16 ms, the adaptive step size was 2 and there were 2 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 2 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 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., 1996). 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.

2. Results

2.1 Word recognition and phoneme identification for natural recordings

Figure 2 displays the range 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. 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.

All of these measures were significantly correlated, $p < 0.01$. For entire words, $r = 0.75$ for non-target and initial target words, $r = 0.71$ for non-target and final target words, and $r = 0.87$ for initial and final target words. For phonemes, $r = 0.73$ for syllables and initial target phonemes, $r = 0.50$ for syllables and final target phonemes, and $r = 0.72$ for initial and final target phonemes.

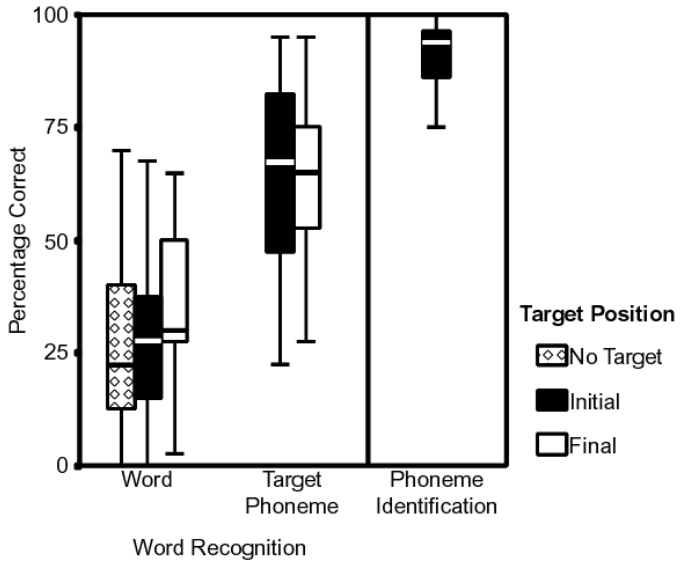

***Figure 2**: Boxplots of results for the open-set word recognition and phoneme identification tasks using natural recordings of speech. Non-target words were those that did not have either /t/ or /d/. Target-initial words had either /t/ or /d/ in syllable initial position. Target-final words had either /t/ or /d/ in syllable final position.*

2.2 Phoneme identification and discrimination: Synthetic stimulus continuum

2.2.1 Sensitivity Functions

Figure 3 displays *sensitivity functions* (i.e., DL values along the VOT continuum) and identification boundary locations. The normal-hearing subjects were relatively homogenous. 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 for normal-hearing subjects were thus generally consistent with previous work on categorical perception (Liberman et al., 1957; Studdert-Kennedy et al., 1970; Repp, 1984).

The sensitivity functions from cochlear implant subjects were highly variable, to an extent that would make the presentation of group sensitivity functions meaningless. Instead, the sensitivity functions from 3 example cochlear implant subjects are displayed in Figure 3. Subject 1 is an example of a cochlear implant user who had results that were consistent with categorical perception. 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 was inconsistent with categorical perception. 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 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.

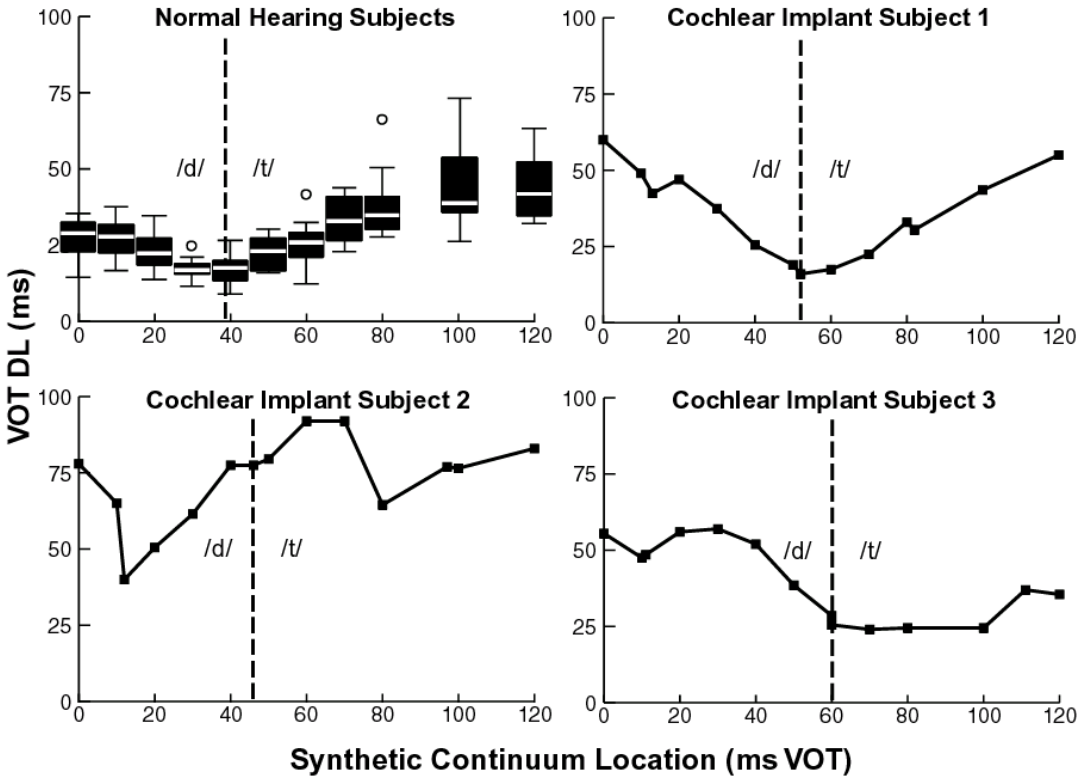

Figure 3: *Boxplots of sensitivity functions for normal-hearing subjects and individual sensitivity functions for 3 example cochlear implant subjects. The vertical dashed lines indicate the location of the phoneme identification boundary. The normal-hearing subjects were fairly homogenous and had results consistent with categorical perception (i.e., high sensitivity at the category boundary, low sensitivity within phoneme categories). The data from the cochlear implant subjects was highly variable; there is evidence of categorical perception for Subject 1, but not for Subjects 2 and 3.*

The adaptive discrimination procedure may not have been valid for Subject 2. For instance, the sensitivity peaks could have occurred because the subject made unsystematic responses (although there was no clear evidence for this from inspection of the raw data), or because there was a non-monotonic relationship between VOT difference and sensitivity (e.g., there may have been local regions of high sensitivity for specific stimuli). Despite this unusual sensitivity function, the identification boundary for this individual was similar to those found for normal-hearing subjects.

Subject 3 is an example of an intermediate case. Within, the /d/ category, the sensitivity function was consistent with categorical perception; sensitivity was poor within the /d/ category and increased near the category boundary. However, 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.

2.2.2 Location Measures: Sensitivity Peaks and Identification Boundaries

To further examine whether cochlear implant users had results that were consistent with categorical perception, the sensitivity peak and category boundary locations were compared. As displayed in Figure 4, 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.

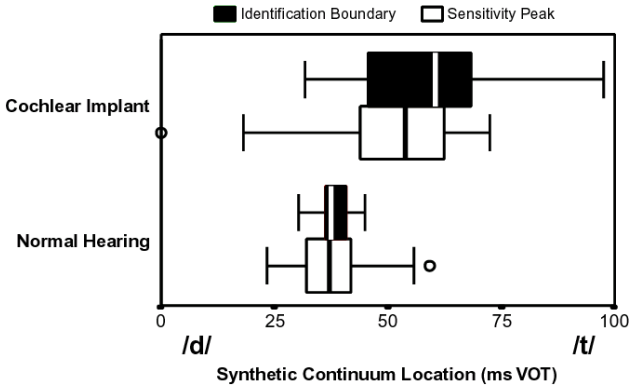

Figure 4: *Boxplots of the locations of identification boundaries and sensitivity peaks along the synthesized stimulus continuum. The distributions of both measures were shifted to longer VOT values for cochlear implant users, compared to those of normal-hearing individuals. Moreover, the individual differences were greater for cochlear implant users than for normal-hearing individuals, on both location measures.*

Figure 5 displays a scatterplot comparing the locations of sensitivity peaks and category boundaries for cochlear implant subjects. The correlation between these two location measures was significant, $r = 0.49$, $p < 0.01$. However, it is also true that few cochlear implant users had their sensitivity peak and identification boundary at exactly the same location (i.e., few points fell on the diagonal in Figure 5). 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.

2.2.3 Sensitivity measures: Minimum DLs and identification boundary widths

As displayed in Figure 6, 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 near identification boundaries and sensitivity peaks. These two measures were correlated for cochlear implant users, $r = 0.47$, $p < 0.01$.

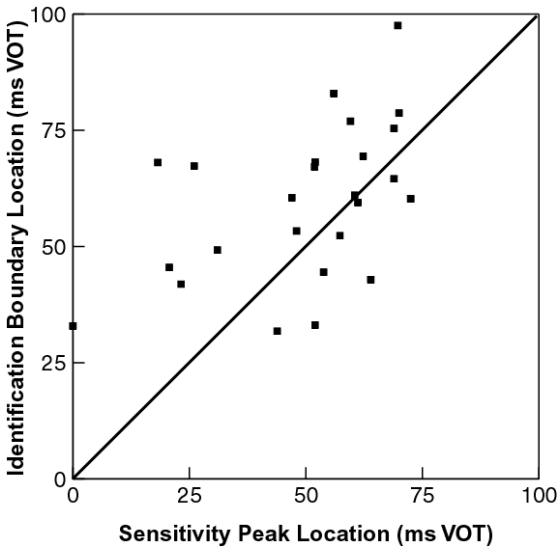

Figure 5: *Scatterplot of the identification boundary and sensitivity peak locations, for cochlear implant users; the solid line is the relationship between the two locations predicted by categorical perception (i.e., identification boundaries and sensitivity peaks at the same locations). Although there was a significant relationship between these two location measures, there was also a substantial amount of scatter.*

2.3 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. For example, in Figure 7 there is a strong relationship between the sensitivity peak locations and the recognition of the initial target phoneme within words. Specifically, subjects who had sensitivity peaks near 45-50 ms most accurately recognized the phoneme targets, and phoneme recognition declined for subjects whose' sensitivity peaks were shifted to shorter or longer VOT values. The relationship was much weaker between the identification boundary locations and the recognition of initial target phonemes within words, but again there was a tendency for individuals with poor phoneme recognition accuracy to have identification boundaries at relatively short or long VOT values.

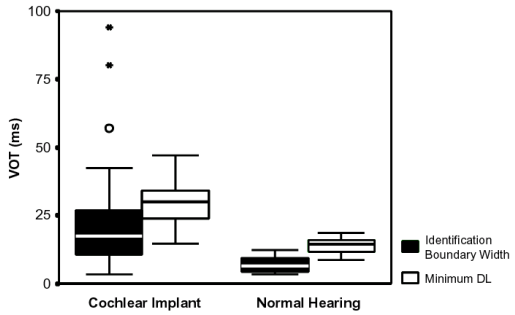

Figure 6: *Boxplots of identification boundary widths and minimum DL values. Although there is overlap between the distributions for individuals with cochlear implants and normal hearing, the cochlear implant users generally had greater widths and DL minima, demonstrating that they were less sensitive to VOT differences near their /d/-/t/ phoneme boundary.*

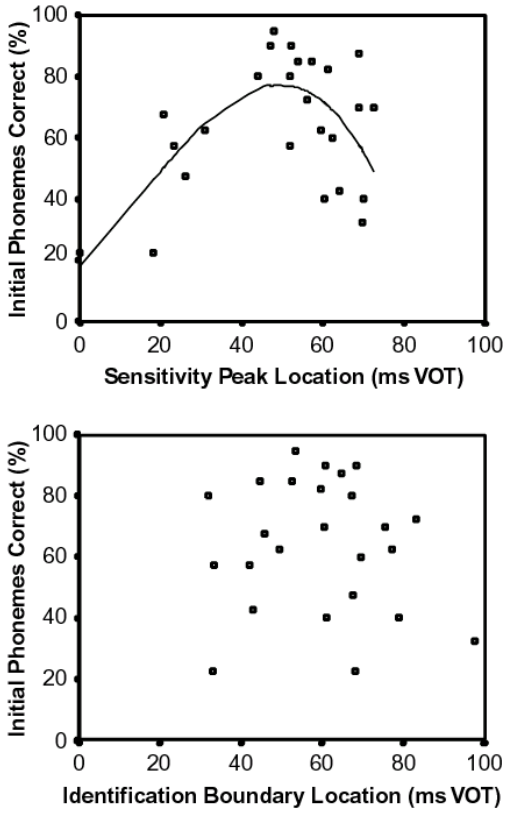

Figure 7: *Scatterplots of the relationships between sensitivity peak and boundary locations with the percentage correct target phoneme recognition within words. There was a significant curvilinear relationship between sensitivity peak location and phoneme recognition. The cubic regression line displayed on the graph had* R^2 *= 0.424, F(21) = 5.16, p = 0.008, demonstrating that subjects with a sensitivity peak at a 45-50-ms VOT tended to have higher phoneme recognition scores. There was only weak evidence of a similar curvilinear relationship between identification boundary locations and phoneme recognition.*

Because of the inverted u-shaped relationship described above, the location measures were re-calculated in terms of their distance from the optimal location along the stimulus continuum (i.e., the point, 47.5 ms, that was related to the highest levels of speech recognition performance). Pearson correlation coefficients between these phonetic perception and speech recognition measures are displayed in Table 1. The results demonstrate 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 were significant.

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 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 (identification boundary width and the difference between the identification boundary and sensitivity peak locations were significantly correlated, $r = 0.48$, $p < 0.01$). The minimum DL is more strongly related to the overall sensitivity to acoustic differences along the series (when correlated with the average DL for each subject, $r = 0.44$ for identification boundary width and $r = 0.84$ for minimum DL; the difference between these correlations was significant, $z = -2.34$, $p < 0.05$).

	Words			Phonemes in Words		Forced-Choice Identification
	Non-target	Initial	Final	Initial	Final	
Optimaility of identification boundary location	-0.27	-0.26	-0.21	-0.31	-0.09	-0.24
Optimality of sensitivity peak location	-0.38*	-0.47*	-0.47*	-0.70*	-0.45*	-0.53*
Identification boundary width	-0.32	-0.29	0.45*	-0.28	-0.37*	-0.16
Minimum DL	-0.29	-0.16	-0.10	-0.06	-0.18	-0.11

* p < .05

Table 1: *Correlation (r) Of Measures of Word Recognition and Phonetic Perception for Cochlear Implant Subjects*

To further test the contribution of all of the phonetic measures to word recognition accuracy, an ANCOVA 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 Figure 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 poor levels of sensitivity.

3. 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 inter-subject 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.

At least from inspection of the scatterplots, the optimum location for a sensitivity peak appeared to be at a longer VOT location (48 ms) than the average normal-hearing sensitivity peak location (37 ms). There is not enough data in this study, particularly for sensitivity peaks at relatively short VOTs, to determine whether this difference is reliable. However, it is plausible that this optimum sensitivity peak location resulted from the ranges of VOTs in the natural stimuli used in this study; VOTs were as long as 51 ms for /d/ and as short as 48 ms for /t/, so a sensitivity peak location in the range of 48-51 ms may have been best for this stimulus set.

It was surprising that word recognition accuracy was unrelated to the 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 of speech perception by cochlear implant users; it seems more important for listeners to have relatively high sensitivity to critical phonetic differences than it is for listeners to be more sensitive to acoustic differences overall.

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 known to be 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.

Perhaps a more likely hypothesis is that cochlear implant users differ in their use of acoustic cues. The articulatory specification of the synthetic continuum resulted in acoustic cues that varied naturally, but not uniformly. For example, stimuli with very short VOTs varied in terms of F1 frequency at the onset of voicing, but stimuli with longer VOTs had similar F1 onsets (the initial rapid rise in F1 was completed within 10 ms of the closure release). Listeners who were more sensitive to F1 onset may thus have been more likely to have sensitivity peaks at shorter VOT locations. This hypothesis raises yet another question: What would cause cochlear implant users to differ in their use of acoustic cues? It is plausible that this variability in cue-weightings also occurs among normal-hearing individuals (Hazan and Rosen, 1991), and that these individual differences, and the functional importance of these differences, are magnified when the available phonetic information is reduced. It is also plausible that the individual differences arise from changes to speech recognition strategies following prolonged periods of deafness and the subsequent accommodation to electric hearing.

This study provides clear evidence that at least some of the perceptual phenomena associated with the categorical perception of consonants have functional importance for the recognition of natural spoken words. Previous cross-language evidence has pointed in this direction. 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). However, the causality of these cross-language results are difficult to interpret because lexical representations and phonetic perception are *both* dependent on language experience. The relationship is clearer in the present study, because cochlear implants directly affect the incoming phonetic information, and the lexical representations of these postlingually deafened subjects were presumably normal. Furthermore, it is unlikely that the observed sensitivity functions were caused by higher-level categorization processes (i.e., by the assignment of category labels), at least for the majority of subjects, because the sensitivity peaks and identification boundaries were often at different locations.

Finally, the finding that the shape of the sensitivity function is particularly important for word recognition is in accord with current developmental accounts of normal-hearing speech perception. Speech perception has been shown to be altered by linguistic experience prior to the age at which word meanings are thought to be acquired (Kuhl et al, 1992), and it has been hypothesized (Kuhl, 1994, 2000; Kuhl and Iverson, 1995) that this early tuning of phonetic perception facilitates the later

acquisition of linguistic categories by making critical phonetic differences more salient. The present results suggest that perceptual tunings can also affect word recognition even after higher-level linguistic categories have been acquired. If perception, or the speech stimulus itself, is altered such that within-category phonetic differences become more salient than between-category differences, then speech recognition performance will be impaired.

Acknowledgements
This work was supported by research grants 1 R03 DC03999 and 2 P50 CD 00242 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health; and grant 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.

References

Bradlow, A. R., & 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., & 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., & 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.

Chatterjee, M. (1999). "Temporal mechanisms underlying recovery from forward masking in multielectrode-implant listeners," J Acoust Soc Am 105, 1853-1863.

Connine, C. M., Blasko, D. G., & 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., & 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., Hannley, M. T., McCandless, G. A., & Smith, L. M. (1988). "Auditory/phonetic categorization with the Symbion multichannel cochlear implant," J Acoust Soc Am 84, 501-510.

Dorman, M. F., Loizou, P. C., Kemp, L. L., & 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., & 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., & 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.

Hanekom, J. J., & Shannon, R. V. (1998). "Gap detection as a measure of electrode interaction in cochlear implants," J Acoust Soc Am 104, 2372-2384.

Hazan, V., & Rosen, S. (1991). "Individual variability in the perception of cues to place contrasts in initial stops," Percept Psychophys 49, 187-200.

Hedrick, M. S., & 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). Sommerville, MA: Sensimetics Corporation.

Hochmair-Desoyer, I. J., Hochmair, E. S., & Stiglbrunner, H. K. (1985). Psychoacoustic temporal processing and speech understanding in cochlear implant patients. In R. A. Schindler & M. M. Merzenich (Eds.), *Cochlear Implants* (pp. 291-304). New York: Raven Press.

Iverson, P. & Kuhl, P. K. (1995). "Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling," J Acoust Soc Am, 49, 467-471, 97, 553-562.

Iverson, P. & Kuhl, P. K. (1996). "Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/," J Acoust Soc Am, 92, 1130-1140.

Johnson, K., & Mullennix, J. W. (1997). *Talker variability in speech processing* (San Diego: Academic Press), Ed. 1. ed.

Kuhl, P. K. (1994). "Learning and representation in speech and language," Curr Opin Neurobiol 4, 812-822.

Kuhl, P. K. (2000). "A new view of language acquisition," Proc Natl Acad Sci U S A 97, 11850-11857.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 121-154). Timonium, MD: York.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). "Linguistic experience alters phonetic perception in infants by 6 months of age," Science 255, 606-608.

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," J Acoust Soc Am 49, 467-471.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). "The discrimination of speech sounds within and across phoneme boundaries," Journal of Experimental Psychology 54, 358-368.

Luce, P. A., & Pisoni, D. B. (1998). "Recognizing spoken words: the neighborhood activation model," Ear Hear 19, 1-36.

Naito, Y., Hirano, S., Honjo, I., Okazawa, H., Ishizu, K., Takahashi, H., Fujiki, N., Shiomi, Y., Yonekura, Y., & Konishi, J. (1997). "Sound-induced activation of auditory cortices in cochlear implant users with post- and prelingual deafness demonstrated by positron emission tomography," Acta Otolaryngol 117, 490-496.

Norris, D. G., McQueen, J. M., & Cutler, A. (2000). "Merging information in speech recognition: Feedback is never necessary," Behavioral and Brain Sciences 23, 299-325.

Parkinson, A. J., Parkinson, W. S., Tyler, R. S., Lowder, M. W., & 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., & 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. (2000). "Cognitive factors and cochlear implants: some thoughts on perception, learning, and memory in speech perception," Ear Hear 21, 70-78.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. H. (1992). *Numerical recipes in C : the art of scientific computing* (Cambridge ; New York: Cambridge University Press), 2nd ed.

Repp, B. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and Language* (Vol. 10, pp. 243-335). New York: Academic.

Rubinstein, J. T., Wilson, B. S., Finley, C. C., & 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 R. S. Tyler (Ed.), *Cochlear implants : audiological foundations* (pp. 357-388). San Diego, Calif.: Singular Pub. Group.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). "Speech recognition with primarily temporal cues," Science 270, 303-304.

Sommers, M. S., Kirk, K. I., & Pisoni, D. B. (1997). "Some considerations in evaluating spoken word recognition by normal- hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format," Ear Hear 18, 89-99.

Stevens, K. N., & Bickley, C. A. (1991). "Constraints among parameters simplify control of Klatt formant synthesizer," Journal of Phonetics 19, 161-174.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & 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.

The Celex Lexical Database. (1995). (Version 2.5). Nijmegen: Center for Lexical Information, Max Planck Institute for Psycholinguistics.

Tong, Y. C., Busby, P. A., & Clark, G. M. (1988). "Perceptual studies on cochlear implant patients with early onset of profound hearing impairment prior to normal development of auditory, speech, and language skills," J Acoust Soc Am 84, 951-962.

Tyler, R. S., Moore, B. C., & Kuk, F. K. (1989). "Performance of some of the better cochlear-implant patients," J Speech Hear Res 32, 887-911.

Waltzman, S. B., Cohen, N. L., & Shapiro, W. H. (1992). "Use of a multichannel cochlear implant in the congenitally and prelingually deaf population," Laryngoscope 102, 395-399.

Yamada, R. A. (1995). Age and acquisition of second language speech sounds: perception of American English /ɹ/ and /l/ by Native Speakers of Japanese. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 305-320). Timonium, MD: York.