On the Structure of Phoneme Categories in Listeners With Cochlear Implants

Harlan Lane  
Massachusetts Institute of Technology, Cambridge, and Northeastern University, Boston, MA

Margaret Denny  
Massachusetts Institute of Technology

Frank H. Guenther  
Boston University, Boston, MA, and Massachusetts Institute of Technology

Helen M. Hanson  
Massachusetts Institute of Technology

Nicole Marrone  
Boston University and Massachusetts Institute of Technology

Melanie L. Matthies  
Boston University and Massachusetts Institute of Technology

Joseph S. Perkell  
Massachusetts Institute of Technology and Boston University

Ellen Stockmann  
Massachusetts Institute of Technology

Mark Tiede  
Massachusetts Institute of Technology and Haskins Laboratories, New Haven, CT

Jennell Vick  
University of Washington, Seattle, and Massachusetts Institute of Technology

Majid Zandipour  
Massachusetts Institute of Technology and Boston University

Phoneme labeling, discrimination and, more rarely, goodness ratings have been reported for several special populations, among them children with apraxic disorders (Maassen, Groenen, & Crul, 2003), children with otitis media (Groenen, Crul, Maassen, & van Bon, 1996), and adults with age-induced sensorineural hearing loss (Ginzel, Brahe, Spliid, & Andersen, 1982; Stevenson, 1975).

The present study is concerned with examining phoneme labeling, discrimination, and category structure in a group of late-deafened adult cochlear implant users. These listeners are expected to be limited in their ability to perceive certain phonemic contrasts using their implants in the first place because of the many differences between electrical and auditory stimulation (Dorman, Loizou, Spahr, & Maloff, 2002). For example, implants stimulate locations more basal than those stimulated directly by the same sound; thus, the listener must learn a new place-frequency code (Rosen, Faulkner, & Wilkinson, 1999; Shannon, 2002). Then, too, implant users’ ability to discriminate formant frequencies is limited by aspects of the frequency-to-electrode map of their implant speech processors (Harmsberger et al., 2001), including spectral resolution (Fu, Shannon, & Wang, 1998), spectral contrast (Loizou & Poroy, 2001), and temporal coding (Xu, Thompson, & Pfingst, 2005). In the second place, the implant user may be hindered by any changes that have taken place in
the auditory–perceptual system as a result of hearing impairment and then prosthesis use.

In an early experiment relating vowel recognition to formant perception, White (1983) studied multiple-choice recognition errors and similarity judgments of pairs of synthesized vowels in a participant by using a 16-electrode implant and a single-channel analog processor. Both perceptual measures showed that the participant had greater sensitivity to first-formant values than to second-formant values. With a second participant, a three-channel processor was also evaluated in addition to the single channel, using an initial consonant test and a vowel test. Consonant voicing and first formant information proved more available with the three-channel processor than with the single-channel processor.

Dorman and Nelson (2005) conducted experiments on phoneme categorizing in noise and in quiet with implant users and normal-hearing listeners. They found that implant users’ phoneme labeling functions were most vulnerable to noise when the phoneme contrast involved rapidly changing spectral patterns. Accordingly, phoneme labeling for the synthesized /ɪ/-/ɔ/ contrast was least affected and semivowel and liquid contrasts were most affected. Nevertheless, implant users had shallower labeling slopes (greater labeling equivocation) for the vowel contrast than did normal-hearing listeners. The processing strategies used by their participants’ implants (and by those of participants in the present study) filter speech signals into frequency bands, extract the amplitude envelope of each band, and send coded pulse trains to the electrodes implanted in the cochlea. When such processing was simulated with four or eight frequency bands, hearing participants’ phoneme labeling functions for the /ɪ/-/ɔ/ contrast formed more distinct categories with the greater number of processing bands.

Dorman, Hanley, McCandless, and Smith (1988) examined labeling of synthesized phoneme contrasts in a listener (subject SS) wearing a four-channel Symbion implant, who had particularly good speech understanding. When labeling stimuli from an /ɪ–ɛ/ series (F1 and F2 covarying), the participant was highly consistent, nearly always assigning a given stimulus value to the same phoneme category, yielding a step function. No step function was in evidence, however, for the /ɪ–ʌ/ series, where F1 was fixed and F2 varied. The authors concluded that when F1 is well resolved, poor resolution of F2 is sufficient for normal identification of front vowels. When the participant was presented with stimuli from a high-frequency voiceless contrast (/s–ʃ/), his labeling function was step-like and within the range of normal performance (see also Dorman, Dankowski, McCandless, Parkin, & Smith, 1991).

Iverson (2003) recruited 25 postlingually deaf participants who as a group used a variety of implant devices and processing strategies. They were asked to label and to discriminate stimuli drawn from a synthesized voice onset time (VOT) opposition /d–t/. Both tasks yield sensitivity measures: steeper labeling slopes indicate greater sensitivity between categories, while smaller difference limens indicate greater within-category discrimination (cf. Munson & Nelson, 2005). The study found larger identification boundary widths, which covary withslope, and larger minimum difference limen values compared with hearing controls. Iverson suggests that performance on vowel and on VOT contrasts may both reflect the implant user’s ability to process spectral cues (e.g., F1 transitions).

In addition to changes in phoneme labeling and discrimination, another alteration in speech processing that may have taken place due to changes in the auditory perceptual system and prosthesis use concerns the internal structure of perceptual phoneme categories. There is a substantial literature showing that listeners with normal hearing are not indifferent to variations within phoneme categories but rather judge some as exceptionally good instances; that is, phoneme prototypes (Iverson & Kuhl, 1995; Miller, 1994). Whether implant users manifest internal category structure to the same degree has not been investigated previously. Moreover, there is evidence that experience plays a role in shaping phoneme categories and their internal structure (Kluender, Lotto, Holt, & Bloedel, 1998), so if implant users present anomalous processing of phoneme contrasts during early implant use, they may improve with time.

In summary, prior studies provide some evidence that implant users are less categorical in their labeling of phoneme contrasts than are speakers with normal hearing; however, those studies used early prostheses (except Iverson, 2003) and did not assess the effects of experience with the implant. Only one study has addressed implant users’ difference limens, which was for a single continuum, VOT (/d–t/). Finally, there are no studies of which we are aware of the effects of implant use on the internal structure of phoneme categories. The present study addresses these issues—labeling, discrimination, goodness rating, and the effects of implant use—with two contrasts: a vowel contrast /a–i/ and a sibilant contrast /s–ʃ/. We chose these two spectral contrasts because our research had shown that implant use enhances their production (Matthies, Svirsky, Perkell, & Lane, 1996; Perkell, Lane, Svirsky, & Webster, 1992), and that the extent of produced contrast covaried across speakers with accuracy in discriminating the contrast (Perkell, Guenther, et al., 2004; Perkell, Matthies, et al., 2004). Moreover, Shannon (2002) emphasized the relative importance of spectral information for speech recognition over amplitude and temporal cues.

These considerations motivate the following two hypotheses:

**Hypothesis 1.** Implant users’ phoneme labeling, discrimination, and within-category structure will all be
anomalous compared with controls: (a) category boundaries will be less well defined; (b) discrimination within categories will show poorer auditory sensitivity; and (c) phoneme prototypes will be less salient, with shallower goodness rating slopes and smaller separation between contrasting prototypes.

Hypothesis 2. Implant users’ phoneme labeling and within-category structure, while anomalous early in implant use, will improve with 1 year’s experience of prosthetic hearing.

Method

Participants

The experimental group comprised 5 male and 2 female postlingually deaf, adult, paid volunteers (Table 1) who were native speakers of English. The group was heterogeneous in several respects. Age at onset of profound loss ranged from 26 to 72. Duration of profound loss varied from less than 1 year to 10 years; the average age of receiving an implant ranged from 28 to 78. The implant was either the Clarion (Advanced Bionics; Kessler, 1999) or the Nucleus 24 device (Cochlear Corporation; Blamey, Dowell, Brown, Clark, & Seligman, 1987; McKay & McDermott, 1993). The implant users were referred to our laboratory by the Massachusetts Eye and Ear Infirmary or the University of Massachusetts Memorial Medical Center.

A control group was made up initially of 3 male and 7 female participants, who reported no speech or hearing anomalies. A screening test was administered to the 8 speakers over age 40, in order to determine approximate thresholds at 0.5, 1, 2, and 4 kHz. After a practice tone at 50 dB HL, sound pressure was increased in 5-dB increments from 0 to 25 dB HL. The series was presented twice at each of the four frequencies to each ear. Participants who failed to report hearing the tone in any of the 16 series (two ears, four frequencies, two trials) were excluded; this resulted in disqualifying 2 male participants. The final control group comprised 1 male and 7 female participants, mean age 45. On a test of phoneme recognition, each hearing participant scored at least 96% correct in vowel recognition (group mean = 98%) and 90% correct in consonant recognition (group mean = 96%).

Stimuli

Two types of stimuli were presented at a comfortable listening level via Harman/Kardon speakers (Harman Multimedia) in a sound-attenuating room.

Natural stimuli. Eight vowel stimuli in /pVT/ context (“pat, pet, pete, pit, poot, pot, put, putt”) and 11 consonants

Table 1. Characteristics of participants with cochlear implants.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto-immune</td>
<td>19</td>
<td>5</td>
<td>Early 30s</td>
<td>18</td>
<td>10</td>
<td>20</td>
<td>Birth</td>
</tr>
<tr>
<td>Infection</td>
<td>54</td>
<td>45</td>
<td>43</td>
<td>28</td>
<td>48</td>
<td>72</td>
<td>26</td>
</tr>
<tr>
<td>Hereditary</td>
<td>56</td>
<td>46</td>
<td>49</td>
<td>28</td>
<td>48</td>
<td>78</td>
<td>36</td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Infection</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>(WW II) Noise</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Hereditary</td>
<td>3.6</td>
<td>3.8</td>
<td>3.6</td>
<td>3.8</td>
<td>3.6</td>
<td>3.8</td>
<td>3.6</td>
</tr>
<tr>
<td>Age at onset of change in hearing (in years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at onset of profound loss (in years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at cochlear implantation (in years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of profound deafness (in years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hearing aid used pre-CI: left, right, both</td>
<td>None</td>
<td>None</td>
<td>Left</td>
<td>None</td>
<td>None</td>
<td>Both</td>
<td>Both</td>
</tr>
<tr>
<td>Implant</td>
<td>Nucleus-24</td>
<td>Clarion C1</td>
<td>Clarion C1</td>
<td>Clarion Auria</td>
<td>Nucleus-24</td>
<td>Clarion C1</td>
<td>Clarion C1</td>
</tr>
<tr>
<td>HiFocus</td>
<td>HiFocus</td>
<td>HiFocus</td>
<td>HiFocus</td>
<td>HiFocus</td>
<td>HiFocus</td>
<td>HiFocus</td>
<td>HiFocus</td>
</tr>
<tr>
<td>Strategy</td>
<td>Spectral peak coding</td>
<td>Simultaneous analog stimulation</td>
<td>High-resolution sequential stimulation</td>
<td>High-resolution sequential stimulation</td>
<td>Advanced combination encoders</td>
<td>Continuous interleaved sampling</td>
<td>Continuous interleaved sampling</td>
</tr>
</tbody>
</table>

Note. CI = cochlear implants. For initials listed at the top of column heads (e.g., F. I., M. J.), the first letter represents “male” or “female,” and the second letter represents the order of the participant’s recruitment, indicated by letter of alphabet.
in /Cot/ context (“bot, cot, dot, got, lot, pot, rot, shot, rot, zot”) were produced by a male and a female speaker for a test of phoneme recognition. Participants listened to at least six repetitions of each of three productions of each of the 19 syllable types. Vowel and consonant syllables were tested separately. Each test comprised three blocks, with stimuli arranged in quasi-random order within a block. The vowel test included 215 trials; the consonant test included 198 trials (for more detail, see Lane et al., 2007).

**Synthetic stimuli.** Because it was desired to have stimuli that sounded as much like natural human speech as possible, copy synthesis was used to create “boot”–“beet” and “said”–“shed” continua (see Hansen, 1995, for details). Recordings were made of 1 male and 1 female participant saying “a boot,” “a beet,” “a said,” and “a shed.” The participants were asked to read the phrases as naturally as possible. Each of the four phrases was read three times, and the 12 utterances were arranged in random order. These recordings were made to digital audiotape in a sound-attenuated chamber at a sampling rate of 48,000 samples/s. After transfer to a computer, the recordings of “a boot” and “a beet” were down sampled to 11,000 samples/s, and the recordings of “a said” and “a shed” were down sampled to 16,000 samples/s. One token of each of the four phrases (per speaker) was chosen as the basis of the copy synthesis. These choices were based on factors such as formant transitions, durations, amplitudes, and regularity of voicing.1

**Vowel continuum.** In the first stage of the synthesis procedure, vowels were synthesized to match the vowel portions of the natural utterances “boot” and “beet.” A Klatt formant synthesizer was used (Klatt, 1980; Klatt & Klatt, 1990). Parameters that varied with time included formants F1–F4, their associated bandwidths, F0, and the voice–source parameters amplitude of voicing, amplitude of aspiration, and spectral tilt. These parameters were adjusted until the synthesized vowels sounded identical to the natural vowels, and the spectra and spectrograms of the synthesized vowels closely matched those of the natural utterance. In this way, the synthetic vowels have the formant–frequency variation and transitions that naturally occur over the course of a syllable.

In the second stage of vowel synthesis, the synthesizer parameters for the vowels were adjusted to create endpoint stimuli that differed only in their F2 and F3 trajectories. This adjustment involved changing the F0, F1, F4–F6, B1–B6, AV, AH, and TL trajectories to have the same values and durations for both /u/ and /i/. In general, these adjustments were minor, as expected. (For example, because both /u/ and /i/ are high vowels, there is little difference in the F1 trajectory.) The F2 and F3 trajectories were adjusted to have the same duration as the other parameters, but otherwise retained their original values. Informal listening tests verified that the stimuli sounded natural, just as though they were spoken by the original speakers.

In the third stage of vowel synthesis, a continuum was created between /u/ and /i/ by linearly interpolating the F2 and F3 trajectories. Seven intermediate stimuli were created. The increment in F2 from the ith to the jth stimulus was given by

\[ |F2_i(t) - F2_j(t)|/8 \tag{1} \]

and likewise for F3 (see Figure 1A in the Appendix). Formant transitions at vowel onset and offset were intermediate between those of the endpoints /u/ and /i/. In addition to the seven intermediate stimuli, two additional stimuli were added at each end of the continuum to extend beyond the parameter values of /u/ and /i/. The same increment, given in Equation 1, was used for these stimuli as was used for the intermediate stimuli.

In the final stage of the vowel synthesis, an initial /b/ (from closure to voice onset, including release) and a final /t/ (from closure to shortly past the release) were excised from one of the natural tokens of “beet” and concatenated with the 13 stimuli of the vowel continuum to create the “boot”–“beet” continuum. (The vowels were in a stop environment, and thus there were no discontinuities in the waveforms produced.) This method was simpler than synthesizing the /b/ and /t/ segments, and informal listening tests verified that the stimuli did indeed sound natural. The trajectories for F2 and F3 of the /a/-/æ/ continua appear in Figure A1.

**Sibilant continuum.** The usual approach for synthesizing a /s/-/ʃ/ continuum is to vary formant frequencies; for example, to shift F3 from a value typical for /ʃ/ to a higher value that is typical of the lowest friction-excited resonance of /s/ (Nittrouer, 1992). However, the acoustic contrast between /s/ and /ʃ/ is due less to a contrast in formant frequencies than to one in formant amplitudes (Keyser & Stevens, 2006). The contrast in formant amplitudes results from a difference in the formants that are affiliated with the front cavity (and thus are strongly excited by friction) and the back cavity (weakly excited). Acoustic theory (Stevens, 1998) predicts that the lowest front-cavity resonance for /ʃ/ will be F3, while the lowest front-cavity resonance for /s/ will be F4 or F5; observations of natural data, including those recorded for this study, bear that out. Thus, the method we used to synthesize our /s/-/ʃ/ continuum was to vary formant amplitudes rather than formant frequencies.

In the first stage of sibilant synthesis, segments were synthesized to match the sibilant portions of the natural

---

1Because of lack of space and the complexity of the synthesis procedure, we do not provide complete synthesis parameters in this article. However, text files of the synthesizer parameters are available at http://hdl.handle.net/1721.1/29217. Readers can access audio files of the stimuli at the same Web site.
utterances “said” and “shed.” The Klatt formant synthesizer was used again; because the sound source was the frication source, the synthesizer was used in parallel mode. In that mode, formant amplitude is controlled by parameters A2F–A6F, which set the amplitudes of the formants. (In the cascade mode, typically used for vowels, formant amplitude is largely controlled by formant frequency and bandwidth.) The synthesized sibilants /s/ and /ʃ/ were matched to the natural sibilants by adjusting the formant frequency and amplitude values until the experimenter judged that the power spectrum of the synthesized sibilant was a good match to that of the natural sibilant. The synthesizer parameters did not vary with time because little formant variation was observed in the natural sibilants.

In the second stage of sibilant synthesis, the synthesizer parameters for /s/ and /ʃ/ were adjusted to create endpoint stimuli that differed only in A2F–A6F. This adjustment involved changing F2–F6 to have the same values for both sibilants. Only minor adjustments were necessary, and informal listening tests confirmed that the stimuli sounded natural and of good quality.

In the third stage of sibilant synthesis, a continuum was created between /s/ and /ʃ/ by interpolating the amplitudes of F2–F6, that is, the parameters A2F–A6F. Seven intermediate stimuli were created. The step size between stimuli was not linear, for two reasons. First, the synthesizer required integer values for the parameters; therefore, a strictly linear interpolation was not possible. Second, pilot tests showed that normal-hearing participants perceived more than half the stimuli as /s/; therefore, adjustments were made in step size to result in normal-hearing participants hearing half the stimuli as /s/ and half as /ʃ/. In addition to the seven intermediate stimuli, two additional stimuli were added at each end of the continuum to extend beyond the parameter values of /s/ and /ʃ/. For some of the formants, the step sizes used for these additional stimuli were significantly larger than those used for the intermediate stimuli, in order for them to sound less than optimal to normal-hearing participants. The formant frequency and amplitude values that were used are summarized in Table A1.

In the last stage of sibilant synthesis, a final /ed/ (from voice onset of the /e/ to shortly past the release of the /d/) was excised from one of the natural tokens of “said” and “shed,” and that single segment was concatenated with all of the 13 stimuli of the sibilant continuum to create the “said”–“shed” continuum. This concatenation was straightforward because, for the two speakers we used as models, /s/ and /ʃ/ did not differ greatly in their transitions into the following vowel. Furthermore, there is usually a transition region with very low amplitude between unvoiced sibilants and a following vowel, so there were no issues of waveform discontinuities. In addition, the use of the same formant frequencies for both /s/ and /ʃ/ simplified the matchup between the synthesized sibilants and the natural vowel. Informal listening tests verified that the stimuli did indeed sound natural.

In our experiments on the role of implant users’ speech perception in their speech production (Perkell, Guenther, et al., 2004; Perkell, Matthies, et al., 2004), we have presented stimuli derived from a female speaker to female participants and stimuli derived from a male speaker to the male participants. We did this in order to approximate the conditions of self-hearing associated with implant users’ produced contrasts. We found that the accuracy of gender-matched discrimination of a phoneme contrast covaried with the extent of that contrast when produced (Perkell, Guenther, et al., 2004; Perkell, Matthies, et al., 2004). We have continued the practice of gender matching in the labeling, discrimination, and goodness rating tasks in the present study.

Participant Tasks

Seven cochlear implant users served individually in sessions conducted approximately 1 month after activation of their implant speech processors, and then again approximately 1 year after activation. One month was judged as close as practicable to processor activation and initial tuning; 1 year was judged sufficiently long after activation to observe effects of prosthesis use, based on our prior studies with vowels (Perkell et al., 1992) and sibilants (Matthies et al., 1996). Only one time sample was elicited from the normal-hearing controls.

Tasks With Natural Stimuli

The natural stimuli described earlier were presented in a phoneme recognition test to implant users before activation of their speech processors, again at 1-month postactivation, and, finally, at 1-year postactivation. The normal-hearing speakers also took the test once. The details of method and results have been given elsewhere (Lane et al., 2007). Briefly, listeners were asked to choose the syllable they heard by clicking on an orthographic version displayed on a computer monitor; the program then proceeded to present the next stimulus. Implant candidates who wore hearing aids preimplant used their aids during the phoneme recognition task in the session prior to implant activation, but in the sessions at 1-month and 1-year postactivation they used only their cochlear implants. Their vowel recognition scores averaged, pre-implant, 36% (SD = 14%) and 1-year postimplant 75% (SD = 16%). Consonant recognition scores were, respectively, pre 27% (SD = 9.5%) and post 68% (SD = 17.8%).

Hearing and hearing-impaired participants also labeled, discriminated, and rated synthesized stimuli. The tasks were administered with ECOS/Win software.
(AVAAZ Innovations). Because the tasks in this experiment were accompanied in each experimental session by other tasks for other experiments, they were often spread out over more than one session. In the 1-month time sample, implant users typically completed these three tasks and phoneme recognition in two sessions, 4 days apart; in the 1-year sample, the tasks were completed in two sessions, 3 days apart.

**Labeling**

During labeling of the members of a contrast, the participant used a computer monitor and mouse to register whether the stimulus presented was “boot” or “beet” when labeling stimuli from the vowel continuum or “said” or “shed” when labeling the sibilant stimuli. In each labeling session, the participant heard 12 repetitions of each of the stimuli in a continuum arranged in random order (using ECOS software); the presentation of the jth stimulus awaited the participant’s response to the ith stimulus. Normal ogives were fit by least squares to each participant’s labeling function, and the slope of the function was determined in the region of 50%. The stimulus number closest to the 50% identification was labeled the category boundary.

**Discrimination**

An ABX paradigm was used to measure implant users’ within-category discrimination after 1 year’s experience with their prostheses (Liberman, Harris, Kinney, & Lane, 1961). The first two stimuli in each triad differed by one, two, or three steps on the synthesis continuum. There were 12 one-step comparisons (1–2, 2–3, ... 12–13), 11 two-step comparisons, and 10 three-step comparisons. Each of these 33 comparisons was presented in four sequences; to illustrate, using Stimuli 1 and 3: 131 133 311 313. The interstimulus interval was 1 s, and the program awaited the listener’s response before presenting the next triad. After each triad was presented, the participant used the computer mouse to indicate whether the third stimulus was the same as the first (A) or the second (B) in the triad. For the illustrative example just cited, the correct responses would be, respectively, A, B, B, and A. There were 12 repetitions of each triad in quasi-random order with the restriction, to reduce context effects, that the last stimulus in a triad was never the same as the first stimulus in the following triad. The vowel and sibilant discrimination tasks were conducted in separate sessions. To obtain a measure of within-category discrimination, the ABX scores of each participant (excluding the triads involving that participant’s category boundary stimulus) were converted to d’ following the procedure of Macmillan and Creelman (1991). Then these values of d’ were pooled over one-, two-, and three-step comparisons and over listeners in each group, implant users, and controls.

**Goodness Rating**

In the goodness rating task, some of the stimuli presented for the labeling task were presented again. To lessen the testing burden on participants (because each session gathered findings for several experiments), when a participant was asked to rate stimuli as exemplars of a given phoneme, the algorithm presented only the stimuli he or she had previously labeled as that phoneme. Some stimuli had been inconsistently labeled; for example, the same stimulus might have been labeled “boot” seven times and “beet” five times. In that case, that stimulus would be presented seven times when the participant was rating items as “boot” and five times when the participant was rating items as “beet.” The participant rated the goodness of each stimulus as a member of its labeling category using a 5-point scale: Responses ranged from 5 (excellent), 4 (good), 3 (fair), 2 (bad), to 1 (terrible). The slopes of the goodness rating functions were determined by linear regression and then multiplied by 13/5: the ratio of the ranges of the x and y variables. Category separation was computed for each participant as the distance between the goodness rating peaks in the vowel contrast and, separately, in the sibilant contrast.

Because the dependent variables of slope, d’, and category separation were frequently nonnormal in their distribution, nonparametric tests were used to test within-group (Wilcoxon signed rank z) and between-groups (Mann–Whitney U) hypotheses (Siegel & Castellan, 1988).

**Results**

Figures 1 and 2, for the vowel and sibilant contrasts, respectively, present the results of the three tasks for the two groups of participants. In each figure, results for implant users are on the left, controls on the right. The plots in the top rows show labeling functions (percentage identification scale at the right). For the implant users, labeling results are shown from two sessions: those at 1 month (open circles) and those at 1 year (filled squares, dotted line). The solid bars in these plots report the average accuracy of within-category ABX discrimination expressed as d’ (left-hand scale). The average pools over both categories in a phoneme contrast, three step sizes, and 7 listeners. The second row of plots shows the goodness rating functions obtained from the implant users at
1 month and from the controls (excellent coded as “5”). Finally, the implant users’ average goodness rating functions at 1 year after activation of their speech processors are shown in the last row.

**Labeling**

We consider first the labeling functions (in the top row of the figures). Comparing implant users to normal-hearing controls (Table 2), the former have shallower slopes as expected both in the 1-month sample of vowel labeling and in the 1-year sample. This outcome is reflected in the average labeling functions plotted in the top row of Figure 1. Six of the seven implant users yielded vowel labeling slopes at 1 month that were shallower than any of the corresponding slopes from the control participants. At 1 year, implant users’ labeling boundaries were more sharply demarcated: Six of the 7 implant users had steeper labeling slopes after 1 year of prosthesis use than they had after 1 month of prosthesis use (Wilcoxon $z = 1.90, p < .05$); and 4 listeners in this group yielded vowel labeling slopes within the range of those of the controls.
Table 2. Average slopes of synthetic speech labeling functions for a vowel and a sibilant contrast.

<table>
<thead>
<tr>
<th>Group</th>
<th>Vowel</th>
<th>Sibilant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>−6.5</td>
<td>−6.9</td>
</tr>
<tr>
<td>Implant users</td>
<td>−2.2</td>
<td>−3.9</td>
</tr>
<tr>
<td></td>
<td>48*</td>
<td>42*</td>
</tr>
<tr>
<td>U</td>
<td>48*</td>
<td>42*</td>
</tr>
</tbody>
</table>

Note. Data from 7 control participants in one time period and 7 implant users in two time periods. U = Mann–Whitney test of group differences. *p < .01.

For the sibilant labeling functions (Figure 2 top row, Table 2), the differences in the slopes of the average labeling functions of implant users and controls were not statistically reliable (p = .05 in the 1-month sample, p > .05 in the 1-year sample). At the 1-month sample, 4 of the implant users yielded sibilant labeling slopes that were shallower than any of the corresponding slopes from control participants. At 1 year, however, all of the implant users yielded sibilant labeling slopes that fell within the range of those obtained from controls. This is the result of the implant users substantially increasing their sibilant labeling slopes following 1 year of implant use; all 7 had steeper slopes at 1 year than they had at 1-month postactivation (Wilcoxon z = 2.40, p < .01); compare the labeling functions at 1 year (Figure 2, closed squares, dashed lines) and at 1 month (open circles, solid lines). Consistent with this outcome, Dorman et al. (1988), as mentioned in the introduction, found a steplike labeling function for the /s–z/ contrast with slope in the range of normal for a listener using a four-channel implant.

Discrimination

Within category discrimination, d′ is plotted by the bars in Figures 1 and 2. It will be seen that, even after 1 year of prosthesis use, synthetic speech stimuli were less discriminable for implant users than for controls. Mean values of d′ for the vowels were 0.7 for implant users and 1.8 for controls (Mann–Whitney U = 5,936, p < .001); for the sibilants, mean values of d′ were 0.9 for implant users and 1.5 for controls (U = 9,549, p < .001).

Goodness Rating

Goodness rating functions appear in the second and third rows of Figures 1 and 2. Each point in the plots is the average goodness rating assigned to the indicated stimulus by all speakers in the group pooled. Table 3 reports the means of the individual participants’ slopes for vowels and sibilants in two time periods for the implant users and one for the controls. For the vowel /i/ and both sibilants, goodness rating functions had more shallow slopes for the implant users than for the controls, indicating they were less selective in the range of acceptable phoneme prototypes. However, there was no difference in implant user and control slopes for /u/, and the difference for /s/ at 1 year postactivation was not statistically reliable.

Although most individual rating functions resembled the corresponding average functions plotted in Figures 1 and 2, there was variation among listeners in both shape and slope of their functions. For most implant users and some controls, the stimuli with the highest ratings were at the two extremes of the continuum presented. The slopes of individuals’ rating functions varied widely. Nevertheless, there was limited overlap in the steepness of implant users’ and controls’ goodness rating functions when the difference in mean slopes was reliable. Comparing steepness of implant users’ rating functions for “beet” to those of controls, there was a single case of overlap between the two groups at 1 month and two cases at 1 year; for “said,” the overlap was one case in the 1-month sample, and for “shed,” the overlap was two cases in both time samples.

At the 1-year time sample, although the average slopes of the implant users’ goodness rating functions for “beet” and “said” became somewhat steeper, this was not the case for “boot” and “shed”; neither of these changes was statistically reliable. For “said” at 1 month and for both sibilants at 1 year, there is evidence of a prototype peak. Recall that during sibilant synthesis, the step sizes used for the additional stimuli at the ends of the continuum were significantly larger than those used for the intermediate stimuli, in order for them to sound less than optimal to normal-hearing participants. That measure had some success.

Table 3. Average slopes of goodness rating functions for synthetic speech stimuli.

<table>
<thead>
<tr>
<th>Group</th>
<th>Boot</th>
<th>Beet</th>
<th>Said</th>
<th>Shed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>−0.65</td>
<td>1.50</td>
<td>−0.94</td>
<td>1.60</td>
</tr>
<tr>
<td>U</td>
<td>29*</td>
<td>29*</td>
<td>1*</td>
<td>5*</td>
</tr>
</tbody>
</table>

Note. Data from 7 control participants in one time period and 7 implant users in two time periods. U = Mann–Whitney Test of Group Differences. *p < .01.
Table 4. Average separation of peak-rated members of two synthesized phoneme contrasts.

<table>
<thead>
<tr>
<th>Group</th>
<th>Vowels</th>
<th>Sibilants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Month</td>
<td>Year</td>
</tr>
<tr>
<td>Controls</td>
<td>9.5</td>
<td>11.3</td>
</tr>
<tr>
<td>Implant users</td>
<td>9.7</td>
<td>10.7</td>
</tr>
<tr>
<td>U</td>
<td>29</td>
<td>35</td>
</tr>
</tbody>
</table>

Note. Data from 7 hearing controls and 7 implant users, and a test of significance between groups (Mann–Whitney U). Ratings were collected from implant users in two time periods but one from controls.

*p < .01.

The obtained goodness rating functions were also parameterized by a measure of phoneme category separation; namely, the number of stimulus units separating the peak ratings of the stimuli in each phoneme contrast. Table 4 reports mean values and the outcome of a test of their significance. Consistent with the average functions shown in the second row of Figures 1 and 2, the separation of the sibilant peaks (Figure 2) was 11.3 stimulus units for the controls, but only 7 stimulus units for the implant users at 1 month, and 8 at 1 year. The differences between the groups are statistically reliable, as indicated in Table 4. However, there were no significant differences between controls and implant users in peak separation for the vowel contrast. The implant users’ increases in peak separation from 1 month to 1 year, shown in Table 4, were not statistically reliable for the vowel contrast (Wilcoxon signed ranks, *z* = 1.63, *p* = .10) nor for the sibilant contrast (*z* = 1.70, *p* = .10).

**Correlation With Phoneme Recognition Scores**

As stated earlier, the implant users in the present experiment also served in a study of phoneme recognition; their results have been reported previously (Lane et al., 2007). Briefly told, after 1 year of prosthesis use, vowel and consonant recognition had improved significantly over preimplant baseline. Speakers who had higher sibilant recognition scores in that study at 1 year tended to have steeper (more categorical) labeling functions in the present experiment: *r* = −.85 (*n* = 7, *p* < .05; recall that the labeling function plots the percentage of “said” responses and thus has a negative slope). The corresponding sibilant correlation at 1 month and those for the vowel stimuli in both time samples were not significant, but it should be kept in mind that the number of implant users was few. Thus, the statistical power was low for correlational analyses. In the study of VOT perception by cochlear implant users cited earlier, Iverson (2003) found, as did the present study, that “individuals with a broader phoneme identification boundary tended to have more difficulty recognizing these phonemes within natural speech” (p. 1062). Iverson did not find a significant correlation between the minimum difference limen for VOT and phoneme recognition, and in the present study no correlation was found between the *d’* measures for vowel and sibilant contrasts and phoneme recognition.

**Discussion**

Because of implant users’ hearing impairment prior to receiving an implant, and because of the differences between electrical and auditory stimulation, we hypothesized that phoneme labeling, discrimination, and within-category structure would all be anomalous for implant users compared with controls. In particular, Hypothesis 1 predicted first that category boundaries in phonemic identification would be less well defined among implant users. The average slopes of the vowel labeling functions were indeed flatter for the implant users than for the normal-hearing controls (Table 2). This was true for sibilant labeling at 1 month and not true at 1 year, but neither between-groups comparison for the sibilants was statistically significant.

Sensitivity within categories was expected to be poorer among implant users. This group proved to have *d’* values approximately half those of the controls: 39% for the vowels, 60% for the consonants (see bar plots embedded in Figures 1 and 2). The identification task requires retaining the stimulus’ auditory properties in memory long enough to make a comparison to an internal representation. However, the ABX task confronting the listener in the present study is more complex. It can be analyzed into two successive tasks. The listener must hold the A stimulus in memory in order to compare it with the B stimulus, and then decide if the order is AB or BA. When the third stimulus arrives, the listener must decide if it is a match to the most recent stimulus. If so, the correct answer is 2; if not, 1 (Macmillan & Creelman, 1991). To what extent lower *d’* scores reflect degraded spectral and temporal properties of the stimuli due to the implant and to what extent there is degraded short-term, echoic memory (Crowder & Roediger, 2001) remains to be investigated.

Hypothesis 1 stated in its third prediction that implant users’ goodness rating functions would be shallower than controls. In fact, all of the mean implant users’ goodness rating functions reported in Table 3 had the same or lesser slope than the corresponding functions of the controls; three of the comparisons, however, were not statistically reliable. Finally, Hypothesis 1 predicted that
the goodness rating functions of implant users would be marked by smaller separation between contrasting prototypes compared with controls. This proved to be true of the bilabials but not of the vowels. The range of prototypes associated with each phoneme was not uncovered in this study: Most of the individual and group goodness rating functions did not show a clear peak, presumably because extending the synthetic continuum two additional stimuli beyond the natural endpoints did not go far enough to induce lower goodness ratings.

The second hypothesis addressed learning: Would implant users’ labeling and goodness rating functions become more like those of the controls after 1 year of prosthesis use? As it turned out, 6 of the 7 implant users had steeper vowel labeling slopes, and all 7 steeper bilabial slopes after 1 year of prosthesis use and the shapes of their average functions were more similar to those of the controls (Figures 1 and 2). However, no significant changes were found when comparing slopes of goodness rating functions obtained in the first and second time samples. Likewise, the separation between peak-rated stimuli in the vowel contrast and in the bilabial contrast did not improve after 1 year’s prosthesis use. It may be that a year is insufficient time to fully rebuild internal category structure. When Fu, Shannon, and Galvin (2002) introduced a 3-mm shift in the frequency-to-electrode map of Nucleus-22 implant users, speech recognition fell initially and did not recover its original level after 3 months of daily use. Lane et al. (2007) found that after 1 year of prosthesis use, produced vowel and bilabial contrasts were considerably less than those of controls and even less than speakers’ preimplant values when measured with self-hearing masked during their productions.

Several properties of the cochlear implants and their speech processing strategies used by participants in this experiment may have contributed to their impaired performance compared with normal-hearing controls. The spectral properties of speech contrasts are transformed by the implants into place of electrical stimulation in the cochlea, and temporal cues are coded in the waveform of the envelope that modulates pulse trains to the electrodes. Relevant to the vowel and fricative place contrasts in the present study, Xu et al. (2005) reported that availability of spectral cues dominates vowel and consonant place recognition. For vowel recognition, spectral resolution trades off with spectral contrast (Loizou & Poray, 2001): When spectral resolution is poor, a larger spectral contrast is needed for vowel identification and conversely. However, duration (a vowel cue) and formant transition contours (a place cue for the fricative contrast) may also be distorted (Teoh, Neuburger, & Svirsky, 2003). The latter investigators also found that the electrical discharge patterns of fricatives (with the SPEAK coding strategy) had shorter durations of formant transition than their acoustic counterparts. Such spectral and temporal differences between acoustic and electrical stimulation may lead implant users to adopt different cue hierarchies compared with normal-hearing listeners. Consequently, the two groups would typically be expected to label, discriminate, and rate synthetic speech stimuli differently.

Summary

Measures of phoneme labeling, discrimination, and category structure were obtained, using synthesized vowel and sibilant contrasts, from a heterogeneous group of postlingually deafened adult implant users in order to examine the extent to which their phoneme perception was altered by a history of hearing impairment and prosthetic hearing. Comparing individual labeling slopes across participant groups, implant users’ vowel labeling slopes were substantially shallower than those of controls. This was also true for sibilant labeling at 1 month, but by 1 year the implant users’ average slope had surpassed that of the controls. However, the sibilant differences were not reliable.

Within-subject comparisons revealed significant increases in labeling slopes after 1 year of prosthesis use. Implant users’ sensitivity to phonetic differences within phoneme categories was roughly half that of the controls for both the vowels and the sibilants. The average slopes of individual implant users’ goodness rating functions never exceeded those of the controls and were shallower in both time samples for three of the four phonemes rated; they did not improve over 1 year. Prototypes for the sibilant contrast (but not the vowels) were closer to one another on the stimulus continuum compared with controls. Prototype separations for the vowel and for the sibilant contrasts did not improve with 1 year’s prosthesis use. Our participants’ perceptual performances were limited no doubt by the spectral and temporal information delivered to the central nervous system by their prostheses. They may have been limited as well by changes in echoic memory, and in auditory processing strategies more generally, because of their history of hearing impairment.

Acknowledgments

This research was supported by Grant R01-DC003007 from the National Institute on Deafness and Other Communication Disorders. We are grateful to Donald Eddington of the Massachusetts Eye and Ear Infirmary and to Daniel Lee of the Department of Otolaryngology–Head and Neck Surgery, University of Massachusetts Medical School, for referring cochlear implant users to our study. We acknowledge with pleasure helpful advice from Joanne Miller, Department of Psychology, Northeastern University. After the first author’s name, the order of the authors is alphabetical.
References


Appendix (p. 1 of 2).

Table A1. Formant amplitude and frequency values used in the synthesis of the sibilant continua.

<table>
<thead>
<tr>
<th>Female participant</th>
<th>Male participant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A2F</td>
</tr>
<tr>
<td></td>
<td>F2 = 1,900</td>
</tr>
<tr>
<td>1</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>34</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
</tr>
<tr>
<td>5</td>
<td>37</td>
</tr>
<tr>
<td>6</td>
<td>37</td>
</tr>
<tr>
<td>7</td>
<td>38</td>
</tr>
<tr>
<td>8</td>
<td>38</td>
</tr>
<tr>
<td>9</td>
<td>39</td>
</tr>
<tr>
<td>10</td>
<td>39</td>
</tr>
<tr>
<td>11</td>
<td>40</td>
</tr>
<tr>
<td>12</td>
<td>41</td>
</tr>
<tr>
<td>13</td>
<td>42</td>
</tr>
</tbody>
</table>

Note. Formant amplitudes are given in dB and formant frequencies are given in Hz. Stimuli 3 and 11, indicated with bold typeface, are the endpoint stimuli.
Figure A1. Trajectories for F2 and F3 of the /u/ – /i/ continuum. The heavier lines in the figures indicate the trajectories for the endpoint stimuli.