JARO
Journal of the Association for Research in Otolaryngology

Perceptual Adaptation to Spectrally Shifted Vowels: Training with Nonlexical Labels

TIANHAO LI AND QIAN-JIE FU

[1]Department of Biomedical Engineering, University of Southern California, Los Angeles, CA 90089, USA
[2]Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, CA 90057, USA

ABSTRACT

Although normal-hearing (NH) and cochlear implant (CI) listeners are able to adapt to spectrally shifted speech to some degree, auditory training has been shown to provide more complete and/or accelerated adaptation. However, it is unclear whether listeners use auditory and visual feedback to improve discrimination of speech stimuli, or to learn the identity of speech stimuli. The present study investigated the effects of training with lexical and nonlexical labels on NH listeners' perceptual adaptation to spectrally degraded and spectrally shifted vowels. An eight-channel sine wave vocoder was used to simulate CI speech processing. Two degrees of spectral shift (moderate and severe shift) were studied with three training paradigms, including training with lexical labels (i.e., "hayed," "had," "who'd," etc.), training with nonlexical labels (i.e., randomly assigned letters "f," "b," "g," etc.), and repeated testing with lexical labels (i.e., "test-only" paradigm without feedback). All training and testing was conducted over 5 consecutive days, with two to four training exercises per day. Results showed that with the test-only paradigm, lexically labeled vowel recognition significantly improved for moderately shifted vowels; however, there was no significant improvement for severely shifted vowels. Training with nonlexical labels significantly improved the recognition of nonlexically labeled vowels for both shift conditions; however, this improvement failed to generalize to lexically labeled vowel recognition with severely shifted vowels. Training with lexical labels significantly improved lexically labeled vowel recognition with severely shifted vowels. These results suggest that storage and retrieval of speech patterns in the central nervous system is somewhat robust to tonotopic distortion and spectral degradation. Although training with nonlexical labels may improve discrimination of spectrally distorted peripheral patterns, lexically meaningful feedback is needed to identify these peripheral patterns. The results also suggest that training with lexically meaningful feedback may be beneficial to CI users, especially patients with shallow electrode insertion depths.

Keywords: nonlexial label training, perceptual adaptation, spectrally shifted vowels

INTRODUCTION

The cochlear implant (CI) has restored hearing sensation to many profoundly deaf individuals. Most contemporary CIs utilize spectrally based speech processing, in which the input acoustic signal is first divided into a number of frequency analysis bands; the temporal envelope extracted from each analysis band is used to modulate pulse trains of current that are delivered to appropriate electrodes implanted within the cochlea. To recreate the tonotopic distribution of acoustic frequency within the normal cochlea, the temporal envelopes extracted from low-frequency bands are delivered to apical electrodes, and the envelopes extracted from high-frequency bands are delivered to basal electrodes. This acoustic frequency to electrode place mapping provides critical spectral cues for CI users. Because of the

Correspondence to: Tianhao Li • Department of Auditory Implants and Perception • House Ear Institute • 2100 West Third Street, Los Angeles, CA 90057, USA. Telephone: +213-9897521; fax: +213-4130950; email: tianhaol@usc.edu

limited insertion depth (Ketten et al. 1998; Skinner et al. 2002) and/or the location of the implanted electrodes, the input acoustic signal is generally spectrally shifted and spectrally compressed, relative to the normal tonotopic pattern.

The acute effects of spectral mismatch and distortion on speech recognition have been extensively studied in normal-hearing (NH) subjects listening to acoustic CI simulations. Dorman et al. (1997) found that, for NH subjects listening to five-channel sine wave vocoder CI simulations, a 2- to 3-mm upward shift (in terms of basilar membrane distance) between the analysis and carrier bands significantly reduced speech performance. Shannon et al. (1998) found that, for NH subjects listening to acoustic noise-band vocoder CI simulations, a mean basalward shift of 3 mm (across four spectral channels) reduced performance with a four-channel processor to that with a single-channel processor. Fu and Shannon (1999) further showed that spectral shifting degraded speech performance, regardless of spectral resolution (up to 16 channels) for NH subjects.

Acutely measured performance may overestimate the effects of spectral mismatch on speech recognition. Perceptual plasticity, whether due to extended experience or explicit training with spectrally shifted speech, can mitigate some acute deficits in performance. Many recent studies have explored perceptual adaptation to spectrally shifted speech by CI users and NH subjects listening to CI simulations. Fu et al. (2002) found that, in Nucleus-22 CI subjects, shifting the frequency allocation assignment from Table 9 (or 7) to Table 1 (i.e., ~1-octave or 0.68 octave downward shift between the analysis and carrier bands) resulted in an acute deficit in speech performance. However, after 3 months of continuously using Table 1, subjects partially adapted to the spectral shift. In a follow-up study, one of the subjects fully adapted to the spectral shift (Table 7 to Table 1) when the spectral mismatch was gradually introduced (Fu et al., 2005b). Dorman and Ketten (2003) found that, after 1 week of continuous exposure, a CI user could fully adapt to a 3.2-mm basal shift along the basilar membrane, and could partly adapt to a 6.8-mm basal shift.

For NH subjects listening to CI simulations, auditory training has been shown to improve perceptual adaptation to spectrally shifted speech. For example, Rosen et al. (1999) found that NH subjects' recognition of upwardly shifted speech quickly improved after 3 h of exposure, when assisted with audiovisual connected discourse tracking (CDT). Similarly, Fu et al. (2003, 2005a) found that moderate auditory training significantly improved NH subjects' adaptation to spectrally shifted speech. In particular, targeted training with spectrally shifted monosyllabic words significantly improved the NH subjects' recognition of spectrally shifted phonemes.

In these previous training studies, it is difficult to know the source of the acute performance deficit with spectrally shifted speech. For example, it is unclear whether subjects were unable to distinguish the shifted peripheral patterns and were unable to identify the peripheral patterns, or some combination of these processes. In terms of the discriminability of shifted peripheral patterns, many speech cues (e.g., relative formant frequencies, voicing, duration, etc.) are preserved with spectrally shifted speech, albeit delivered to a different cochlear location (Liu and Fu 2006). However, these cues with shifted speech may not be as perceptually robust as compared to tonotopically matched speech. In terms of identification, spectrally shifted peripheral patterns may be in conflict with central speech pattern templates (i.e., abstract knowledge structure of speech stored in the central nervous system) acquired during NH listening experience. Depending on the degree of spectral shift, lexical labels (which refer to these central speech pattern templates) may impose a perceptual bias on response choices. Because both these processes (i.e., discrimination and identification) may contribute to perceptual adaptation, especially with auditory training, it is important to understand their relative contribution. Characterizing the learning transfer is important for understanding the neural basis of perceptual learning (Ahissar 2001). By testing speech recognition performance after training with nonlexical labels, we may better understand the mechanisms that underlie the storage and retrieval of speech patterns in the central nervous system. Better understanding of these mechanisms would help in the design of effective training protocols for postlingually deafened CI users.

In general, postlingually deafened CI users must adapt to the novel peripheral patterns provided by the CI device relative to central speech pattern templates acquired during normal or impaired hearing prior to implantation. Theoretically, auditory training with lexically meaningful feedback will aid the perceptual adaptation process. However, the effect of lexical labels on perceptual adaptation may depend on whether the new peripheral pattern is consistent with the lexical label (and by extension, with the central speech pattern template). Under conditions of severe spectral shifting, an electrically evoked phoneme pattern may coincide with an incorrect lexical label. For example, when severely shifted, the stimuli "heed" and "who'd," while sounding different from each other, may both sound most like "heed." When presented with "who'd," a listener may choose "heed" as the correct response,

as the lexical label "heed" best matches the stimulus. When training or testing with lexical labels for severely shifted vowels, the label may introduce some "cognitive interference" and inhibit the adaptation process. When spectral mismatch is less severe, lexical labels will be less likely to interfere with adaptation and may even facilitate the adaptation process, as the electrical patterns and central speech pattern templates will be more closely aligned. Thus, when speech is severely distorted, it may be preferable to train phoneme recognition using nonlexical labels at the first stage of training, thereby avoiding potential cognitive interference introduced by lexical labels. By using nonlexical labels, listeners may be better able to focus on acoustic difference among phonemes.

In the present study, we studied NH listeners' perceptual adaptation to spectrally shifted vowels by using three training paradigms: training with lexical labels, training with nonlexical labels, and the test-only paradigm. An eight-channel sine wave vocoder was used to generate spectrally shifted vowels; two degrees of spectral mismatch were tested: moderate and severe shift. The results show that with the test-only paradigm, lexically labeled vowel recognition significantly improved for moderately shifted vowels, suggesting that storage and retrieval of speech patterns in the central nervous system is somewhat robust to tonotopic distortion and spectral degradation. Training with nonlexical labels significantly improved the recognition of nonlexically labeled vowels for both shift conditions; however, this improvement failed to generalize to lexically labeled vowel recognition with severely shifted vowels. Training with lexical labels significantly improved lexically labeled vowel recognition with severely shifted vowels. These results suggest that training with lexically meaningful feedback is necessary for CI users, especially patients with shallow electrode insertion depths.

METHODS

Subjects

Sixteen NH subjects (age range: 18–34 years; nine men and seven women) participated in the study. All subjects were native speakers of American English. All subjects had pure-tone thresholds better than 20 dB HL at octave frequencies from 125 to 8000 Hz. Subjects had no prior experience with acoustic CI simulations before the study. All subjects were paid for their participation.

Test and training materials

Vowel stimuli for both training and testing included 12 medial vowels presented in an /h-V-d/ context (i.e., "had," "hod," "hawed," "head," "heard," "hid," "heed," "hood," "hud," "who'd," "hayed," "hoed"). Vowel tokens were digitized natural productions from five male and five female talkers, randomly drawn from the speech samples recorded by Hillenbrand et al. (1995). Thus, the vowel stimulus set for both training and testing contained 120 tokens (12 vowels * 10 talkers).

Signal processing and spectral shift conditions

An eight-channel sine wave vocoder was used to simulate CI speech processing, and was implemented as follows. The input speech signal was band-passed into eight frequency analysis bands (fourth-order Butterworth filters). The temporal envelope was extracted from each analysis band by half-wave rectification and low-pass filtering (fourth-order Butterworth filter with corner frequency at 160 Hz). The temporal envelope from each channel was used to modulate a corresponding sine wave carrier; the sine wave carrier frequencies were varied according to the experimental condition (see below). The modulated sine waves were summed and the output was adjusted to have the same long-term root-mean-square (RMS) energy as the input speech signal (65 dB).

Two degrees of spectral mismatch between the frequency analysis bands and sine wave carriers were investigated: (1) moderate spectral shift and (2) severe spectral shift. For the moderate shift condition, the overall input frequency range was 75–5411 Hz and the overall output frequency range was 150–10,823 Hz, resulting in a 1-octave upward spectral shift. These input and output frequency ranges correspond to frequency allocation Tables 1 and 9, respectively, used in the Nucleus-22 speech processor. The distribution of the analysis and carrier bands was calculated according to Greenwood's (1990) formula, with the assumption that the cochlear is ~35 mm in length. Table 1 shows the analysis frequency range and sine wave carrier frequency for each channel for the moderate shift condition. Note that the degree of spectral mismatch (in terms of cochlear distance) gradually increased from 3.0 mm for the most apical channel to 4.9 mm for the most basal channel, with a mean shift of 4.4 mm toward the base, across all channels. For the severe shift condition, the input overall frequency information range was 200–7000 Hz and the output overall frequency range was 999–10,290 Hz; the output carrier bands were upwardly shifted to simulate a shallow insertion of a 16-mm-long, eight-electrode array with 2-mm electrode spacing. The distribution of the analysis and carrier bands was calculated according to Greenwood's (1990) formula. Table 2 shows the analysis frequency range and sine wave carrier frequency for each channel for the severe shift condition. Note that the

TABLE 1

Corner and center frequencies for the analysis and carrier filters, for the moderate shift condition

Channel number	Analysis band corner frequencies (Hz)	Analysis band center frequencies (Hz)	Greenwood distance from cochlear apex (mm)	Carrier frequencies (Hz)	Greenwood distance from cochlear apex (mm)	Shift distance from apex to base (mm)
1	75–375	194	5.2	368	8.2	3.0
2	375–675	508	10.0	1011	14.1	4.1
3	675–1015	831	12.9	1658	17.3	4.4
4	1015–1539	1253	15.5	2503	20.1	4.6
5	1539–2092	1802	17.9	3583	22.6	4.7
6	2092–2872	2458	20.0	4907	24.8	4.8
7	2872–3942	3358	22.1	6749	27.0	4.9
8	3942–5411	4636	24.4	9263	29.3	4.9

degree of spectral mismatch (in terms of cochlear distance) gradually decreased from 8.3 mm for the most apical channel to 3.1 mm for the most basal channel, with a mean shift of 5.7 mm toward the base, across all channels.

Test and training procedures

For all conditions, vowel recognition was measured by using a 12-alternative forced choice paradigm. A stimulus was randomly selected (without replacement) from the stimulus set and presented to the subject. The subject responded by clicking on one of 12 response boxes, after which a new stimulus was presented. Vowel recognition was measured by using two types of response labels: lexical labels (in which the 12 response boxes were labeled with an /h-V-d/ word, i.e., "heed," "had," "head," etc.) and nonlexical labels (in which the 12 response boxes were simply labeled with a letter, i.e., A, B, C, ..., L). To remove any response bias associated with the position of the response boxes on screen, the vowel stimulus associated with each response box was different between the lexically labeled and nonlexically labeled vowel recognition tests. Figure 1 shows the response box labels and associated vowel stimuli (in

parentheses) for the lexically labeled (panel A) and nonlexically labeled (panel B) vowel recognition tests. During testing, no trial-by-trial feedback was provided, and subjects were instructed to guess if they were not sure of the correct response.

Subjects were randomly assigned to four different groups (four subjects in each group): group 1—training with nonlexical labels (moderately shifted vowels); group 2—test-only with lexical labels, i.e., equivalent exposure to the test stimuli without preview or feedback (moderately shifted vowels); group 3—training with nonlexical labels (severely shifted vowels); group 4—training with lexical labels (severely shifted vowels). The test-only data for severely shifted vowels was taken from Fu et al. (2005a); in that study, four subjects were tested, using identical procedures and speech processing conditions. Note that training with lexical labels for moderately shifted speech was not conducted.

Training and testing were conducted over 5 consecutive days. On the first day, pretraining baseline vowel recognition was measured with unprocessed vowels and with eight-channel, spectrally matched vowels (the center frequencies of the analysis and carrier bands were matched); vowel recognition was measured with lexical labels. The pretraining vowel recognition familiarized subjects

TABLE 2

Corner and center frequencies for the analysis and carrier filters, for the severe shift condition

Channel number	Analysis band corner frequencies (Hz)	Analysis band center frequencies (Hz)	Greenwood distance from cochlear apex (mm)	Carrier frequencies (Hz)	Greenwood distance from cochlear apex (mm)	Shift distance from apex to base (mm)
1	200–359	272	6.7	1168	15	8.3
2	359–591	464	9.4	1585	17	7.6
3	591–930	744	12.2	2136	19	6.8
4	930–1426	1155	14.9	2863	21	6.1
5	1426–2149	1754	17.7	3821	23	5.3
6	2149–3205	2629	20.4	5084	25	4.6
7	3205–4748	3905	23.2	6748	27	3.8
8	4748–7000	5769	25.9	8942	29	3.1

HAD	HOD	HAWED	HEAD
(HAD)	(HOD)	(HAWED)	(HEAD)
HAYED	HEARD	HID	HEED
(HAYED)	(HEARD)	(HID)	(HEED)
HOED	HOOD	HUD	WHO'D
(HOED)	(HOOD)	(HUD)	(WHO'D)

(A)

A	B	C	D
(WHO'D)	(HAYED)	(HEED)	(HOED)
E	F	G	H
(HID)	(HAD)	(HUD)	(HEAD)
I	J	K	L
(HAWED)	(HOD)	(HOOD)	(HEARD)

(B)

FIG. 1. (A) Layout and response boxes for lexically labeled tests and training. This layout was used for the test-only and lexically labeled training paradigms, as well as for all baseline and posttraining recognition tests. The text in parenthesis shows the stimulus associated with the response label. (B) Layout and response boxes for nonlexically labeled tests and training. The text in parenthesis shows the stimulus associated with the response label.

with the test procedure and with the eight-channel vocoder processing. Before training, baseline vowel recognition (with lexical labels) was also measured with moderately shifted vowels (groups 1 and 2) and severely shifted vowels (groups 3 and 4).

Subjects completed two to four training exercises on each training day. Training consisted of a 5-min preview of the vowel test stimuli; vowel recognition was immediately tested following each preview. For training with nonlexical labels (groups 1 and 3), subjects previewed 120 vowel tokens (i.e., 12 vowels produced by 10 talkers) with letter labels; subjects sequentially previewed the 12 vowels for each talker

(i.e., 12 vowels produced by talker 1, followed by 12 vowels produced by talker 2, etc.). After previewing of vowel stimuli with nonlexical labels, vowel recognition with nonlexical labels was immediately tested. Note that pretraining baseline performance with nonlexical labels was not measured, as subjects had no prior knowledge of the association between the labels and the stimuli (consequently, performance would be at chance level, or 8.33% correct). For training with lexical labels (group 4), subjects similarly previewed 120 vowel tokens with lexical labels. After previewing of vowel stimuli with lexical labels, vowel recognition with lexical labels was immediately tested. For all training groups, at the end of the training period, vowel recognition with lexical labels was remeasured for eight-channel spectrally shifted, eight-channel spectrally matched and unprocessed vowels.

RESULTS

Before training was begun, pretraining vowel recognition performance was measured with unprocessed speech and eight-channel, spectrally matched speech. Mean recognition of unprocessed vowels was 91% correct. Mean recognition of eight-channel, spectrally matched vowels was reduced to 77% correct. Note that after training with spectrally shifted vowels, mean performance was not significantly changed for spectrally matched vowels (78% correct). When eight-channel vowels were moderately shifted, mean pretraining performance was further reduced to 45% correct. When eight-channel vowels were severely shifted, mean pretraining vowel recognition was only 9% correct (i.e., nearly chance level performance).

Figure 2 shows recognition performance for moderately shifted vowels, as a function of training or test session. The different symbols show individual subject data for vowel recognition while training with nonlexical labels (i.e., A, B, C, ..., L; group 1); the solid line shows mean performance. The dashed line shows mean performance for vowel recognition for repeated testing with lexical labels (i.e., test-only data; group 2). Subject performance with nonlexical labels significantly improved with training [one-way, repeated-measures (RM) analysis of variance (ANOVA): $F_{(4,12)}=15.389$, $p<0.001$]; mean performance increased correct from 40% to 77% correct. *Post-hoc* Bonferroni *t*-tests showed that performance significantly improved by the second day of training ($p<0.05$), after which performance did not significantly improve. Similarly, performance of subjects in the test-only condition significantly improved as a function of test session [one-way RM ANOVA: $F_{(4,12)}=40.833$, $p<0.001$]; mean performance improved from 44% correct on day 1 to 64% correct on day 5. *Post-hoc* Bonferroni *t*-tests
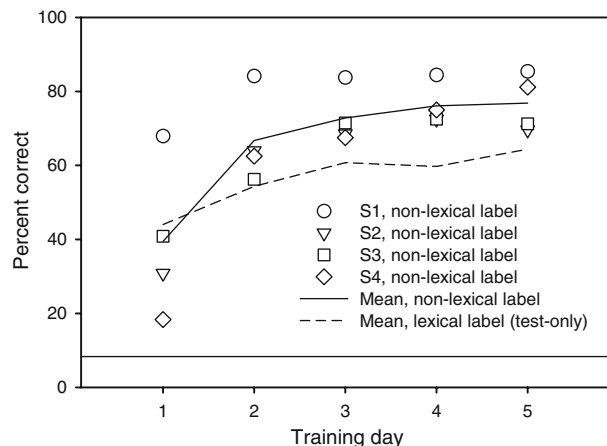
FIG. 2. Percent correct scores for moderately shifted vowels, as a function of training (nonlexical labels) or test (test-only) session. The symbols show individual subjects' vowel recognition performance with nonlexical labels (group 1); the solid line shows mean performance. The dashed line shows mean test-only vowel recognition performance with lexical labels (group 2). The thin solid line shows chance performance level (8.33% correct).

showed that performance significantly improved between the first and second, and between the second and third test sessions ($p<0.05$), after which performance did not significantly improve. At the end of the study period, mean performance with nonlexical labels was ~13 percentage points higher than the test-only performance with lexical labels. A two-way ANOVA showed significant effects for the training paradigm [$F_{(1,30)}=4.680$, $p=0.038$] and for training/test day [$F_{(4,30)}=5.324$, $p=0.002$]; there was no significant interaction between training paradigm and training/test day [$F_{(4,30)}=0.660$, $p=0.625$]. Note that while the training paradigm has a statistically significant effect on performance ($p=0.038$), the analysis did not achieve adequate statistical power (power=0.447), most likely due to the small number of subjects and intersubject variability.

Figure 3 shows individual subjects' vowel recognition performance (group 1) with moderately shifted vowels, before (black bars) and after (white bars) training. The left panel shows recognition performance with nonlexical labels and the right panel shows performance with lexical labels. Note that subjects were trained using nonlexical labels only. Posttraining vowel recognition with both nonlexical and lexical labels was significantly improved after 5 days of training. A two-way RM ANOVA showed a significant effect for training [$F_{(1,3)}=15.135$, $p=0.03$], but not for response label [$F_{(1,3)}=4.605$, $p=0.121$]; there were no significant interactions between training and response labels [$F_{(1,3)}=7.036$, $p=0.077$].

Similar to Figure 2, Figure 4 shows recognition performance for severely shifted vowels, as a function

of training or test session. The different symbols show individual subject data for vowel recognition while training with nonlexical labels (group 3); the solid line shows mean performance. The dashed line shows mean performance for vowel recognition for repeated testing with lexical labels (i.e., test-only), previously reported by Fu et al. (2005a). The thin solid line shows chance level (8.33% correct). Post-training mean performance with nonlexical labels was much lower for severely shifted vowels (35% correct) than for moderately shifted vowels (77% correct). Similarly, for the test-only paradigm, post-training mean performance was much lower for severely shifted vowels (12% correct) than for moderately shifted vowels (64% correct). Subject performance with nonlexical labels significantly improved with training [one-way RM ANOVA: $F_{(4,12)}=11.971$, $p<0.001$]; mean performance increased from 16% to 35% correct. *Post-hoc* Bonferroni *t*-tests showed that performance significantly improved by the third day of training ($p<0.05$), after which performance did not significantly improve. For the test-only data with severely shifted vowels, there was no significant change in performance over the 5-day study period [one-way RM ANOVA: $F_{(4,12)}=1.779$, $p=0.198$]. At the end of the study period, mean performance with nonlexical labels was ~23 percentage points higher than test-only performance with lexical labels; *post-hoc* Bonferroni *t*-tests showed significant difference between the two groups ($p=0.011$).

Similar to Figure 3, Figure 5 shows individual subjects' (group 3) vowel recognition performance with severely shifted vowels, before (black bars) and after (white bars) training. The left panel shows recognition performance with nonlexical labels and
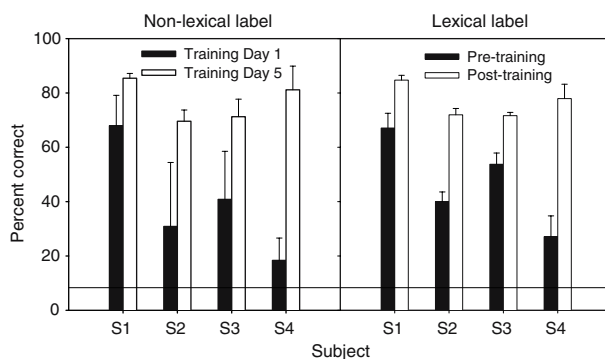

FIG. 3. Individual subject vowel recognition scores for moderately shifted vowels, before (black bars) and after (white bars) training (group 1). The left panel shows vowel recognition performance with nonlexical labels. The right panel shows vowel recognition performance with lexical labels. The error bars represent one standard deviation. The thin solid line shows chance performance level (8.33% correct).
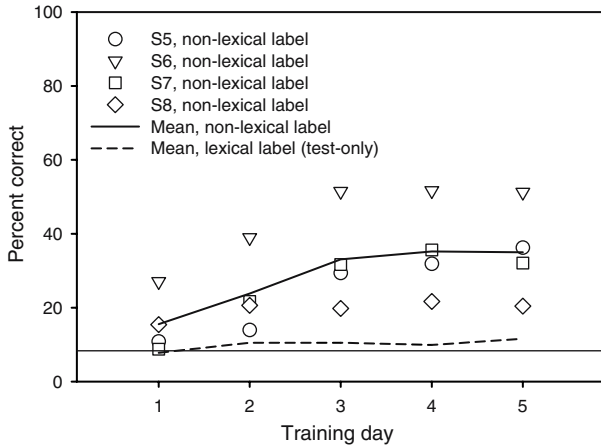
FIG. 4. Percent correct scores for severely shifted vowels, as a function of training (nonlexical labels) or test (test-only) session. The symbols show individual subjects' vowel recognition performance with nonlexical labels (group 3); the solid line shows mean performance. The dashed line shows mean test-only vowel recognition performance with lexical labels (from Fu et al. 2005a). The thin solid line shows chance performance level (8.33% correct).

the right panel shows performance with lexical labels. Again, note that subjects were trained using nonlexical labels only. Although performance with nonlexical labels significantly improved by the end of the training period [one-way RM ANOVA: $F_{(1,3)} = 16.161$, $p = 0.028$], there was no significant improvement in vowel recognition with lexical labels [one-way RM ANOVA: $F_{(1,3)} = 1.086$, $p = 0.374$].

Figure 6 shows mean vowel recognition performance with severely shifted vowels as a function of training and test session for the three training

protocols. The circles show test-only performance with lexical labels (from Fu et al. 2005a). The triangles show performance with lexical labels while training with lexical labels. The squares show performance with nonlexical labels while training with nonlexical labels. The thin solid line shows chance level performance (8.33% correct). As reported in Fu et al. (2005a), there was no change in test-only performance over the 5-day study period. However, performance significantly improved with training, whether subjects were trained and tested with nonlexical or lexical labels [one-way RM ANOVA: $F_{(4,12)} = 10.865$, $p < 0.001$; for group 4]. *Post-hoc* Bonferroni *t*-tests showed that performance did not significantly improve until the fourth day; there was no significant difference in performance between the fourth and fifth days of training. This rate of improvement was slightly slower than that of group 3 (nonlexical label training), which reached asymptotic performance levels by the third day of training. Comparing performance between groups 3 and 4, a two-way ANOVA showed significant effects for training paradigm [$F_{(1,30)} = 6.476$, $p = 0.016$] and training/test day [$F_{(4,30)} = 5.935$, $p = 0.001$]; there no significant interaction between training paradigm and training/test day [$F_{(4,30)} = 0.174$, $p = 0.950$]. Note that although the training paradigm has a statistically significant effect on performance ($p = 0.016$), the analysis did not achieve adequate statistical power (power = 0.619), most likely as a result of the small number of subjects and intersubject variability.
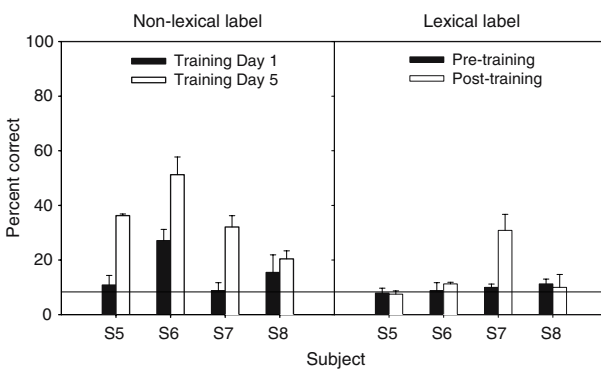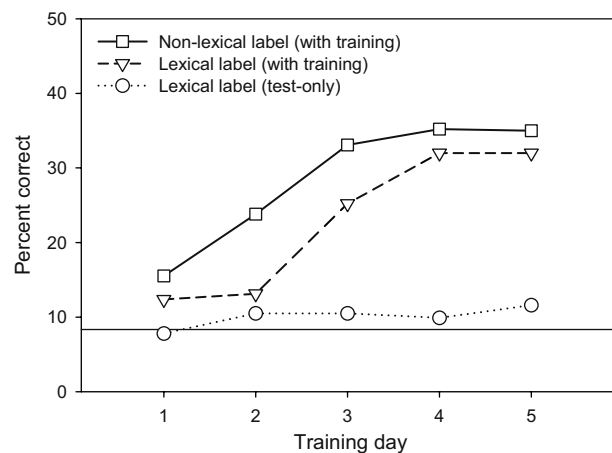

FIG. 5. Individual subject vowel recognition scores for severely shifted vowels, before (black bars) and after (white bars) training (Group 3). The left panel shows vowel recognition performance with nonlexical labels. The right panel shows vowel recognition performance with lexical labels. The error bars represent one standard deviation. The thin solid line shows chance performance level (8.33% correct).

FIG. 6. Mean recognition of severely shifted vowels, as a function of training and test session, for three training paradigms. The circles show test-only performance (from Fu et al. 2005a). The triangles show lexically labeled vowel recognition performance while training with lexical labels (group 4). The squares show nonlexically labeled vowel recognition performance while training with nonlexical labels (group 3). The thin solid line shows chance performance level (8.33% correct).

DISCUSSION

Results from the present study showed that most NH subjects were able to improve recognition of spectrally shifted vowels with auditory training, consistent with previous studies (e.g., Rosen et al. 1999; Fu et al. 2005a). Auditory training improved performance, whether using lexical or nonlexical labels in the training procedure. With a moderate spectral shift, even repeated testing significantly improved recognition performance; however, with a severe spectral shift, there was no improvement in performance after 5 days of repeated testing. When the peripheral pattern is severely distorted, the training protocol (i.e., test-only, training with lexical or nonlexical labels) matters considerably more.

For moderately shifted vowels, performance significantly improved after 5 days of nonlexical label training (group 1) or 5 days of repeated testing (group 2). This result suggests that perceptual adaptation to moderately shifted speech may be mostly stimulus-driven (i.e., without lexically meaningful feedback), and that moderately shifted vowels can be automatically aligned to correct central speech pattern templates. The results further imply that speech storage and retrieval is somewhat robust to tonotopic distortion (up to 1-octave upward shift), even for spectrally degraded speech (eight-channel sine wave vocoder CI simulation). However, at the end of the 5-day study period, mean performance with the nonlexical label training protocol was significantly higher (~13 percentage points) than that with the test-only protocol. The improvement with nonlexical label training also generalized to improved recognition of lexically labeled vowels.

For severely shifted vowels, both overall posttraining performance and the amount of improvement were significantly lower than that with moderately shifted vowels. Moreover, there was no significant improvement in performance with the test-only paradigm. In contrast to moderately shifted vowels, training with nonlexical labels did not generalize to improved recognition of lexically labeled, severely shifted vowels. This result implies that, although training with nonlexical labels may have improved the discriminability among vowel stimuli, severely shifted vowels could not be matched to their central speech pattern templates. It is possible that subjects may have focused on different cues when training with lexical and nonlexical labels. With nonlexical labels, listeners may attend more closely to acoustic differences among stimuli; with lexical labels, listeners may focus on phonetic structure. Analysis of perceptual confusion matrices revealed that training/testing with lexical labels and training/testing with nonlexical labels resulted in improved recogni-

tion of similar vowel stimuli (namely, "hoed," "heed," "hayed," "hid," and "who'd"). This analysis suggests that subjects may have learned similar speech cues with the two training protocols. Because the nonlexical label training did not generalize to improved lexical label testing, the results suggest that severely shifted vowels (although somewhat discriminable) cannot be matched to their central speech pattern templates without lexical feedback. Conversely, the improved performance with lexical label training may have been partly attributable to the trained association between the shifted peripheral patterns and the central speech pattern templates (represented by the lexical labels). By extension, the acute baseline deficit with lexical labels may have been partly a result of conflicts between the peripheral and central speech patterns caused by the lexical labels. With nonlexical labels, this conflict presumably does not exist; subjects had to learn to associate the nonlexical labels with the peripheral patterns.

The different results with moderately and severely shifted vowels may also have been a result of the acoustic properties of the signal after the CI simulation processing. The severely shifted vowels were significantly more spectrally shifted and spectrally compressed than the moderately shifted vowels. The phonetic structure may have been overly distorted by frequency compression, making vowels more difficult to identify. However, results from Baskent and Shannon (2005) showed that, for the basal region of the cochlea (i.e., simulation of a shallow electrode insertion depth), mild amounts of frequency-place compression provided better speech performance than truncating the input frequency range to match the cochlear location. Thus, the degree of frequency shift may have more strongly contributed to the performance deficit. A severe shift may significantly change the perceptual vowel space, reducing the perceptual distance between vowels. For example, when spectrally matched, "who'd and "heed" may sound quite distinct; with a severe shift, they may sound more alike. While the degree of acoustic frequency shifting may contribute more strongly than the amount of acoustic frequency compression, the physiology of the normal cochlea may result in some sort of perceptual frequency compression. For different cochlear locations, the same cochlear extent may produce different amounts of frequency resolution. Although the Greenwood transformation attempts to compensate for these differences, the degree of perceptual frequency compression may be quite different at the base than at the apex of the cochlea. As such, even though the ratio of formant frequencies may have been largely preserved with the Greenwood filter distribution of the acoustic signal, there may have been significant compression of these

ratios when mapped to the basal region of the cochlea. For CI patients, given the uncertainties of auditory neuron health and proximity to electrode locations, the perceptual vowel space may be even more distorted.

It is also possible that the 5-day training period was not sufficient for subjects to adapt to severely shifted vowels. Had subjects been given more time, they may have been able to improve lexically labeled vowel recognition with nonlexical label training. However, the data showed no significant change in performance after the third day of training with nonlexical labels. Also, analysis of the confusion matrices showed that vowels that were easily identified in the nonlexical label tests could not be identified in the lexical label tests. Thus, it seems unlikely that further training with nonlexical labels would have significantly improved lexically labeled vowel recognition.

It is interesting to note that for moderately and severely shifted vowels, training and testing with nonlexical labels provided the best overall performance and largest improvement in performance. With moderately shifted vowels, performance with nonlexical labels (group 1) was significantly better than test-only performance with lexical labels (group 2). With severely shifted vowels, performance with nonlexical labels (group 3) was significantly better than performance with lexical labels (group 4). Although the small number of subjects and inter-subject variability somewhat tempers the significance of these differences in training outcomes, the trends suggest that lexical labels may have introduced some conflict between shifted peripheral patterns and central speech pattern templates. Listeners may be able to better distinguish some phonemes when this conflict is removed. However, because the improved recognition of severely shifted vowels with nonlexical label training did not generalize to improved recognition of lexically labeled vowels, lexically meaningful feedback is needed to associate the shifted patterns with correct central pattern templates.

Results from the present study also have implications for the design of effective auditory training protocols for postlingually CI patients. Because of the limited extent and shallow insertion depth of implanted electrodes, some CI users must adapt to similar degrees of spectral shifting and spectral compression speech as tested in the present study. For CI users with relatively deep insertion depths, the corresponding spectral shift and compression may be small to moderate. These CI users may "automatically" adapt to the distorted electrical speech patterns within a short period, as moderately shifted speech can be automatically aligned with central speech pattern templates (similar to the results for groups 1 and 2). For CI users with shallow insertion depths, the

potential shift and distortion of spectral cues may be quite severe. Some CI users may be able to learn and process the distorted peripheral neural patterns simply as a result of daily exposure; however, this process may be quite long. Gradual introduction of the shift may allow for more complete adaptation (Fu et al. 2005b). Others may benefit from auditory training with lexically meaningful feedback. Such training may incorporate lexical labels or even lip-reading. Alternatively, nonlexical label training may be initially beneficial to improve the discriminability of spectrally shifted speech, after which lexical label training would help to associate the peripheral patterns to the central pattern templates. Results of the present study suggest that the speech performance and training outcomes of CI users might be quite variable, and that training protocols require flexibility to meet the needs of individual CI patients. Lexically meaningful feedback is necessary to build associations between electrically evoked speech patterns and central speech pattern templates, or to reformulate central speech pattern templates to accommodate spectrally distorted peripheral patterns. For prelingually deafened CI patients, lexically labeled training and lexically meaningful feedback are necessary to develop central auditory speech pattern templates, which were not acquired or under-developed during deafness.

CONCLUSION

The present study investigated the effects of auditory training with lexical and nonlexical labels on NH listeners' perceptual adaptation to spectrally shifted and spectrally degraded speech, and compared training outcomes with a test-only paradigm. Two degrees of spectral shift were studied: moderate and severe shift. After 5 days of training with nonlexical labels, recognition of both moderately and severely shifted vowels with nonlexical labels significantly improved. With moderately shifted vowels, performance also significantly improved after 5 days of repeated testing with lexical labels (without preview or feedback). These results suggest that some amount of "automatic learning" of spectrally shifted speech is possible for NH listeners. For severely shifted vowels, nonlexical label training and repeated testing did not significantly improve lexically labeled vowels recognition, suggesting that automatic learning of severely distorted peripheral patterns may not be possible. Interestingly, performance with nonlexical labels was significantly better than that with lexical labels for the severe shift condition, suggesting that there may have been some cognitive conflict between the shifted peripheral patterns and central

speech pattern templates (represented by the lexical labels). Because the nonlexical label training did not generalize to improved recognition of severely shifted speech with lexical labels, the results suggest that training with lexically meaning feedback is necessary to adapt to severely shifted speech; such training may especially benefit CI users with shallow electrode insertion depth.

ACKNOWLEDGMENT

REFERENCES

AHISSAR M. Perceptual training: a tool for both modifying the brain and exploring it. PNAS 98:11842–11843, 2001.

BASKENT D, SHANNON RV. Interactions between cochlear implant electrode insertion depth and frequency-place mapping. J. Acoust. Soc. Am. 117(3):1405–1416, 2005.

DORMAN MF, KETTEN D. Adaptation by a Cochlear-Implant patient to upward shifts in the frequency representation of speech. Ear Hear. 24(5):457–460, 2003.

DORMAN MF, LOIZOU PC, RAINEY D. Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. J. Acoust. Soc. Am. 102(5):2993–2996, 1997.

FU QJ, GALVIN JJ. The effects of short-term training for spectrally mismatched noise-band speech. J. Acoust. Soc. Am. 113(2): 1065–1072, 2003.

FU QJ, SHANNON RV. Recognition of spectrally degraded and frequency shifted vowels in acoustic and electric hearing. J. Acoust. Soc. Am. 105(3):1889–1900, 1999.

FU QJ, SHANNON RV, GALVIN JJ. Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. J. Acoust. Soc. Am. 112(4):1664–1674, 2002.

FU QJ, NOGAKI G, GALVIN JJ. Auditory training with spectrally shifted speech: implications for cochlear implant patient auditory rehabilitation. J. Assoc. Res. Otolaryngol. 6(2):180–189, 2005a.

FU Q-J, NOGAKI G, GALVIN JJ, WANG X. Perceptual learning and auditory training in cochlear implant patients. Abstracts of 2005 Conference on Implantable Auditory Prosthesis, p53, 2005b.

GREENWOOD DD. A cochlear frequency-position function for several species—29 years later. J. Acoust. Soc. Am. 87(6):2592–2605, 1990.

HILLENBRAND J, GETTY LA, CLARK MJ, WHEELER K. Acoustic characteristics of American English vowels. J. Acoust. Soc. Am. 97(5): 3099–3111, 1995.

LIU C, FU Q-J. Estimation of vowel recognition with cochlear implant simulations. IEEE Trans. Biomed. Eng., 2006, in press.

KETTEN DR, VANNIER MW, SKINNER MW, GATES GA, WANG G, NEELY JG. In vivo measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays. Ann. Otol. Rhinol. Laryngol. 175:1–16, 1998.

ROSEN S, FAULKNER A, WILKINSON L. Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. J. Acoust. Soc. Am. 106(6):3629–3636, 1999.

SHANNON RV, ZENG FG, WYGONSKI J. Speech recognition with altered spectral distribution of envelope cues. J. Acoust. Soc. Am. 104(4):2467–2476, 1998.

SKINNER MV, KETTEN DR, HOLDEN LK, HARDING GW, SMITH PG, GATES GA, NEELY JG, KLETZKER GR, BRUNSDEN B, BLOCKER B. CT-derived estimation of cochlear morphology and electrode array position in relation to word recognition in Nucleus-22 recipients. J. Assoc. Res. Otolaryngol. 3(3):332–350, 2002.