Published in final edited form as:

Ear Hear. 2009 December ; 30(6): 662–674. doi:10.1097/AUD.0b013e3181b9c92d.

Transfer of Auditory Perceptual Learning with Spectrally Reduced Speech to Speech and Nonspeech Tasks: Implications for Cochlear Implants

Jeremy L. Loebach¹, David B. Pisoni^{1,2}, and Mario A. Svirsky³

¹Indiana University Department of Psychological and Brain Sciences

²Indiana University School of Medicine

³New York University School of Medicine, Department of Otolaryngology

Abstract

Objective—The objective of this study was to assess whether training on speech processed with an 8-channel noise vocoder to simulate the output of a cochlear implant would produce transfer of auditory perceptual learning to the recognition of non-speech environmental sounds, the identification of speaker gender, and the discrimination of talkers by voice.

Design—Twenty-four normal hearing subjects were trained to transcribe meaningful English sentences processed with a noise vocoder simulation of a cochlear implant. An additional twenty-four subjects served as an untrained control group and transcribed the same sentences in their unprocessed form. All subjects completed pre- and posttest sessions in which they transcribed vocoded sentences to provide an assessment of training efficacy. Transfer of perceptual learning was assessed using a series of closed-set, nonlinguistic tasks: subjects identified talker gender, discriminated the identity of pairs of talkers, and identified ecologically significant environmental sounds from a closed set of alternatives.

Results—Although both groups of subjects showed significant pre- to posttest improvements, subjects who transcribed vocoded sentences during training performed significantly better at posttest than subjects in the control group. Both groups performed equally well on gender identification and talker discrimination. Subjects who received explicit training on the vocoded sentences, however, performed significantly better on environmental sound identification than the untrained subjects. Moreover, across both groups, pretest speech performance, and to a higher degree posttest speech performance, were significantly correlated with environmental sound identification. For both groups, environmental sounds that were characterized as having more salient temporal information were identified more often than environmental sounds that were characterized as having more salient spectral information.

Conclusions—Listeners trained to identify noise-vocoded sentences showed evidence of transfer of perceptual learning to the identification of environmental sounds. In addition, the correlation between environmental sound identification and sentence transcription indicates that subjects who were better able to utilize the degraded acoustic information to identify the environmental sounds were also better able to transcribe the linguistic content of novel sentences. Both trained and untrained groups performed equally well (~75% correct) on the gender identification task, indicating that training did not have an effect on the ability to identify the gender of talkers. Although better than chance, performance on the talker discrimination task was poor overall (~55%), suggesting that either

explicit training is required to reliably discriminate talkers' voices, or that additional information (perhaps spectral in nature) not present in the vocoded speech is required to excel in such tasks. Taken together, the results suggest that while transfer of auditory perceptual learning with spectrally degraded speech does occur, explicit task-specific training may be necessary for tasks that cannot rely on temporal information alone.

Keywords

Acoustic simulations; Cochlear implants; Noise vocoder; Speech perception; Gender identification; Environmental sound identification; Talker discrimination

Introduction

The ability to perceive speech in one's native language is a robust skill for normal-hearing listeners, even under highly degraded listening conditions (e.g. Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Fenn, Nusbaum, & Margoliash, 2003; Pisoni, Manous, & Dedina, 1987; Remez, Rubin, Pisoni, & Carrell, 1981; Schwab, Nusbaum, & Pisoni, 1985; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Normal-hearing listeners' abilities to adapt to spectrally degraded speech develop rapidly after minimal exposure to the processing conditions (e.g. Burkholder, 2005; Davis *et al.*, 2005; Fenn *et al.*, 2003; Remez *et al.*, 1981), are retained over time (Fenn *et al.*, 2003; Roth, Kishon-Rabin, Hildesheimer, & Karni, 2005; Schwab *et al.*, 1985), and generalize to new stimuli or signal processing conditions on which listeners received no training (Burkholder, 2005; Davis, Taylor, Johnsrude, & Carlyon, 2004; Fenn *et al.*, 2003; Fu & Shannon, 1999; Schwab *et al.*, 1985).

Taken together, the speed, retention, and generalizability of auditory perceptual learning in normal-hearing listeners indicate that it is an extremely robust perceptual process. However, the above studies utilized tasks that required the symbolic or linguistic identification of the stimuli. An additional and important benchmark of the robustness of auditory perceptual learning is the ability to transfer and carry over auditory perceptual learning to different nonlinguistic tasks. The present study sought to assess whether training on a linguistic task (the transcription of meaningful sentences processed with a noise vocoder) transfers to the identification of non-speech ecologically-significant environmental sounds, the identification of talker gender, or the discrimination of talkers by their voice alone.

Transfer of perceptual learning has been observed in the auditory (Delhommeau, Michey, Jouvent, & Collet, 2002; Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994), visual (Hunstad, 1985), and motor domains (Murray, 1981; Teixeira, 2000) as well as in amodal cognitive tasks (Benson, Lovett, & Kroeber, 1997; Muramoto, 2001) in both human and nonhuman species (Delay, 2001; Nakagawa, 2000; Watanabe, 1986). Successful transfer of perceptual learning suggests that these processes are extremely flexible, allowing the extension of what was learned under one cognitive or perceptual task to novel untrained tasks. In the auditory domain, transfer of perceptual learning has been demonstrated in both simple psychophysical discrimination tasks (Delhommeau *et al.*, 2002) and in more complex voice recognition and speech perception tasks (Nygaard & Pisoni, 1998; Nygaard *et al.*, 1994). In the experiments of Nygaard and colleagues (1994; 1998), subjects were explicitly trained to identify 5 male and 5 female talkers by voice. Participants were trained with words over a period of 9 days, or with sentences over a period of 3 days, and then tested on the transcription of speech in noise produced by familiar talkers (from training) and unfamiliar talkers. Novel words and sentences produced by familiar talkers (those whom participants were trained to identify by name) were transcribed with significantly higher accuracy than novel words and sentences produced by unfamiliar talkers (Nygaard & Pisoni, 1998; Nygaard *et al.*, 1994). This result suggests that what subjects learned in the talker identification task successfully

transferred to a speech intelligibility task, even though subjects were not instructed to focus on the linguistic content of the stimuli during training.

Studying the transfer of auditory perceptual learning in normal-hearing listeners is of theoretical interest for several reasons. First, a listener's ability to transfer training to new tasks can provide important insights into the underlying cognitive and neural mechanisms that support perceptual learning. Second, examining conditions under which transfer of auditory perceptual learning occurs may help determine what aspects of the auditory signal listeners become attuned to during training. Enhancing our understanding of the transfer of auditory perceptual learning also has important clinical implications for hearing-impaired listeners. Even with appropriate amplification from hearing aids or treatment with a cochlear implant, many hearing-impaired listeners show significant deficits in speech perception and spoken word recognition (Teoh, Pisoni, & Miyamoto, 2004; Teoh, Neuberger & Svirsky, 2003). Methods of aural rehabilitation for hearing aid users have focused primarily on auditory training paradigms that provide the greatest benefit for speech perception, production and language-related skills outside of the clinical setting (Bode & Oyer, 1970; Deguchi, Kagami, & Hiki, 1981; Kennedy & Weener, 1973; Massaro & Light, 2004). The advent of cochlear implants, however, has provided an alternative method for treating profound hearing loss that cannot be successfully treated with amplification alone. The substantial differences between hearing with a hearing aid and hearing with a cochlear implant suggest no apriori reasons to expect that training paradigms that are effective for one group of patients would be effective for the other. Thus, an important new area of research is in the efficacy of training and transfer of perceptual learning in CI users.

The relationship between performance on speech and non-speech auditory perception tasks in CI users has been examined for talker-gender identification (Fu, Chinchilla, & Galvin, 2004), and talker discrimination tasks (Cleary & Pisoni, 2002; Cleary, Pisoni, & Kirk, 2005). Open- and closed-set word and sentence recognition scores in pediatric CI users have been found to be strongly correlated with the ability to discriminate talkers (Cleary & Pisoni, 2002; Cleary *et al.*, 2005). These findings suggest that the children who excel in the word identification and talker discrimination tasks are better able to make use of the spectro-temporal information provided by their implant. The specific relationship between these two abilities, however, has not been conclusively determined for adult CI users (Fu *et al.*, 2004). The few studies that have examined talker discrimination in adult CI users have reported considerable difficulty in discrimination when talkers are the same gender and when there is linguistic variability within the pairs of test stimuli (Kirk, Houston, Pisoni, Sprunger, & Kim-Lee, 2002; McDonald, Kirk, Krueger, & Houston, 2003). Although it has been reported that speech perception skills relate to tasks that do not require accurate encoding of the spoken message (e.g. talker-gender identification, talker discrimination) and to nonspeech auditory identification and discrimination tasks (e.g. music appreciation, sound identification), it is currently unknown whether auditory training on speech will transfer to nonspeech/nonlinguistic tasks.

In addition, there has been only limited research into the abilities of CI users to identify and understand nonspeech stimuli, such as music (Gfeller *et al.*, 2001; Kong, Cruz, Jones, & Zeng, 2004) and environmental sounds (Reed & Delhorne, 2005). Using a closed-set environmental sound identification task, Reed and Delhorne (2005) found that adult CI users could identify a limited number of the environmental stimuli, particularly those sounds that had highly distinctive temporal characteristics (e.g. footsteps, slamming door). Other stimuli with distinctive spectral characteristics (e.g. air conditioner, dishwasher) were identified with significantly less accuracy. In addition, a significant relationship was observed between environmental sound recognition and word recognition: CI users who had good open-set word recognition abilities also identified the more difficult sounds more often than subjects with poor open set word recognition (Reed & Delhorne, 2005). These findings suggest that the

perception of speech and nonspeech sounds may rely on common perceptual processes and cognitive resources.

The perception of environmental sounds is an important skill for profoundly deaf CI users, especially given the link between speech perception and environmental sound identification observed by Reed and Delhorne (2005). In rehabilitative settings with pediatric cochlear implant users, it is not uncommon for clinicians first to teach children the explicit awareness of auditory percepts through sounds in the environment. Children may spend time on “listening walks” that alert them to the sound qualities of birds, cars, or other things in their environment (Robbins, 1998). However, the perception of environmental sounds by adult and pediatric cochlear implant users, and whether these abilities correlate with speech perception abilities, has not been explored in sufficient detail.

Acoustic simulations of cochlear implants using noise vocoders have provided a useful tool for studying many aspects of speech perception (i.e., Dorman, Loizou, & Rainey, 1997; Fu & Shannon, 1999a; Rosen, Faulkner, & Wilkinson, 1999; Shannon *et al.*, 1995) and perceptual learning (Burkholder, 2005; Davis *et al.*, 2005; Fu & Galvin, 2003; Loebach & Pisoni, 2008; Loebach, Bent & Pisoni, 2008; Loebach, Pisoni, & Svirsky, In Press) in normal-hearing listeners. Recently, the investigation of non-linguistic auditory processing skills, such as music perception (Kong *et al.*, 2004; Smith, Delgutte, & Oxenham, 2002), environmental sound identification (Gygi, Kidd, & Watson, 2004; Loebach & Pisoni, 2008; Shafiro, 2008), talker-gender identification and talker discrimination (Gonzalez & Oliver, 2005) have been conducted using the vocoder. However, it is currently unknown whether training with speech processed through acoustic simulations of cochlear implants would result in robust learning that would transfer to music perception, gender identification, or talker discrimination tasks.

Given the links between speech perception and environmental sound identification in CI users (Reed & Delhorne, 2005), it is clinically important to determine whether the perceptual learning of spectrally degraded speech transfers to the identification of environmental sounds. Therefore, the present study was designed to assess the transfer of auditory perceptual learning of spectrally degraded speech to environmental sound and talker-gender identification and talker discrimination in normal-hearing adult listeners trained to transcribe speech processed with an 8-channel noise vocoder. Given that both speech and environmental sound perception occur in response to acoustic simulations of cochlear implants (Gygi *et al.*, 2004; Shannon *et al.*, 1995) and the links between speech perception and environmental sound identification in CI users (Reed & Delhorne, 2005), we hypothesized that training would transfer from a speech transcription task to environmental sound identification. However, given that gender identification and talker discrimination both rely more heavily on the ability to resolve finer acoustic details that may not be well represented in the vocoder, such as resolving and comparing fundamental frequency as well as individual formant frequencies, we did not expect auditory perceptual learning to transfer to talker-gender identification and talker discrimination.

Methods and Materials

Participants

Forty-eight normal-hearing, young, healthy adult participants were recruited from the Indiana University community. All participants were monolingual native speakers of American English, reported being free of any speech, hearing, language, and attentional disorders, and were paid \$12 for their participation in this study.

Stimuli

Six highly familiar nursery rhymes (*Jack and Jill*; Humpty dumpty; Hey diddle diddle; Hot cross buns; Little miss Muffet; Star light, star bright) were used to familiarize the subjects with the acoustic simulation. 120 sentences drawn from lists 11 through 22 of the Harvard sentence corpus (IEEE, 1969) were used as stimuli in pre- and posttest and during training. All stimuli were produced by a female speaker, who had been previously determined to produce highly intelligible speech (Burkholder, 2005). Recordings were made in a sound-attenuated booth (IAC Audiometric Testing Room, Model 402A) using a Shure head-mounted microphone (SM98). Recordings were digitized online (16-bit analog-to-digital converter (DSC Model 240)) at 22,050 Hz, and stored as Windows PCM .wav files. All stimuli were normalized to 65dB(A) RMS.

The sentence materials used in the gender identification and talker discrimination tasks were taken from the Indiana Multitalker Database (Karl & Pisoni, 1994), a collection of 20 male and 20 female talkers each producing the same 100 novel sentences from lists 1 through 10 of the Harvard sentences (IEEE, 1969). Ten novel sentences from each of the five most intelligible male and female talkers (based on previous intelligibility scores, Bradlow, Torretta, & Pisoni, 1996), were selected for use in the gender identification task (100 sentences in total). All sentences were unique and no repetitions were included in the gender identification task. Additionally, sixty of the same sentences were selected from the two most intelligible male and female talkers (240 sentences total) for use in the talker discrimination task, and were combined into 60 same talker and 60 different talker pairs for each gender. All pairings included different sentences for each talker.

Environmental sound identification was assessed using the set of 120 tokens of 40 different sounds compiled by Reed and Delhorne (2003). The 40 sounds were divided into four categories (*General Home, Kitchen, Office, and Outdoors*) according to the environmental context in which the sounds are usually heard (Reed & Delhorne, 2003) based on the criteria of Ballas (1993). The temporal and spectral characteristics of these stimuli were determined previously (Reed and Delhorne, 2005) based on the total signal duration, transiency, ratio of burst duration to total duration, intensity, and frequency region with peak intensity. For a more complete description of the spectral and temporal characteristics of these sounds, please see Reed and Delhorne (2005). Table 1 lists the ten sounds included in each sound category and indicates those identified by Reed and Delhorne (2005) as having distinct transient or temporal properties.

Signal Processing

The signal processing strategy used for the cochlear implant simulation was done using the signal processing methods described in Kaiser and Svirsky (2000). Each stimulus was pre-emphasized using a second order Butterworth low-pass filter (1200 Hz) and divided into eight frequency bands using a bank of band-pass IIR analysis filters. The output of each filter modulated the amplitude of a noise band whose frequency range was identical to that of the corresponding filter. The noise bands were intended to simulate the percepts evoked by different intracochlear electrodes, and their frequency ranges were chosen based on typical CI electrode locations (see Stakhovskaya, Sridhar, Bonham & Leake, 2007): (854-1467 Hz, 1467-2032 Hz, 2032-2732 Hz, 2732-3800 Hz, 3800-5150 Hz, 5150-6622 Hz, 6622-9568 Hz, 9568-11,000 Hz). The amplitude envelope was extracted from each band using a third order Butterworth low-pass filter (150 Hz) and used to modulate bands of white noise that were filtered with the same cutoff frequencies as the original analysis filters. The resulting stimuli contain eight spectral channels that lack the acoustic fine structure of the original stimulus. This is a common way to simulate the perceptual experience of cochlear implant users who

have an electrode array with eight stimulation points (e.g., Kaiser & Svirsky, 2000; Fu *et al.*, 2005).

Procedures

The experimental procedure was divided into five phases: (1) Familiarization, (2) Pretest, (3) Training, (4) Posttest, and (5) Transfer. All training and testing was conducted on the same day, and the entire experiment lasted approximately an hour. Subjects were divided into two groups of 24 subjects. Although both groups transcribed the same Harvard sentences during training, subjects in the control group heard the unprocessed version of the sentences and received no feedback, while subjects in the experimental group heard vocoder processed sentences, and received feedback (in the form of the text of the correct sentence paired with a representation of the degraded stimulus) after they made their responses. Participants were tested in individual testing stations equipped with a Gateway PC (P5-133) using a 15" monitor (Vivitron15). Stimuli were presented over calibrated headphones (Beyer Dynamics DT100) at approximately 70dB(A) SPL.

The experiment began with a brief familiarization phase in which participants listened to and silently read six nursery rhymes that were processed with the vocoder. The text of the rhymes appeared on the monitor 500ms prior to the start of each utterance and remained visible for the duration of the utterance. Listeners did not have the option to replay the stimuli.

The pretest phase immediately followed familiarization. During the pretest phase, listeners transcribed 20 vocoder-processed Harvard sentences by typing their responses on a keyboard immediately after each utterance ended. Subjects did not receive feedback and could not repeat sentences.

Following pretest, subjects began the training phase. Subjects in the experimental group heard 100 Harvard sentences in their processed form, and received feedback in the form of the repetition of the processed sentence while the text of the sentence appeared on the screen (appearing 250ms prior to the onset of each auditory stimulus and remained visible until the sentence stopped playing). Listeners could not replay sentences or the feedback. This feedback was selected based on earlier findings showing that this auditory plus orthographic feedback results in higher pre- to posttest gains compared to auditory-alone feedback (Burkholder, 2005; Davis *et al.*, 2005; Loebach, Pisoni, & Svirsky, In Press). Subjects in the control group heard the same 100 Harvard sentences in their unprocessed form and did not receive any feedback. All subjects transcribed the sentences using the same procedures for pre- and posttest.

The posttest phase followed training, and listeners transcribed the same 20 processed Harvard sentences they heard in the pretest by typing their responses on a keyboard. Responses could be entered immediately after each utterance ended. Participants did not receive any feedback and did not have the option to repeat sentences.

Tests of transfer of perceptual learning followed the posttest, and consisted of three separate tests in which subjects performed various tasks in response to vocoder processed materials: (1) environmental sound identification (2) talker-gender identification, and (3) talker discrimination.

The environmental sound identification task was conducted using a 10-alternative forced-choice procedure (Reed & Delhorne, 2005). The 10 sound alternatives appeared on the computer terminal in numbered order prior to the start of the first sound and remained visible until all 30 tokens from the sound category had been played. Listeners indicated the sound they heard by pressing the numbered key paired with the sound. The order of presentation of the

four sound categories was counterbalanced across listeners. Listeners could not replay any of the sounds.

In the gender identification task, listeners heard 50 Harvard sentences spoken by five different male talkers and 50 Harvard sentences spoken by five different female talkers. Listeners indicated the gender of the talker by pressing the numbered key paired with the gender labels “male” or “female”. Listeners did not have the option to replay sentences.

During the talker discrimination task, subjects heard pairs of sentences and indicated if they were produced by the same or different talker. 120 sentences were paired by talker, so that subjects heard 60 pairings of the same talker (30 male pairs, and 30 female pairs), 30 pairings of different male talkers and 30 pairings of different female talkers. The order of the talker pairings was counterbalanced across all pairings. Each pair was presented with a 1000 ms pause between the first and second sentence. Listeners indicated whether the talkers were the same or different by pressing a numbered key paired with the item “same” or “different”. Listeners could not replay individual sentences or sentence pairs.

Data Analysis

Harvard sentences contain five keywords that subjects must correctly transcribe (e.g. “The beauty of the view stunned the young boy”) (IEEE, 1969). For each sentence, the percentage of keywords correctly transcribed (out of five) was calculated and averaged across each block. Typographical errors were scored as correct if a target letter was substituted by any immediately surrounding letter on the keyboard. Responses in which the correct letters were transposed were also considered as typographical errors and scored as correct. Keywords that contained obvious spelling errors and homophones were also scored as correct. However, changes in the words’ tense or with other incorrect affixes were considered incorrect. Performance was compared across the two groups of subjects (Trained and Untrained).

Performance on the Environmental stimuli was scored as correct or incorrect based on the numerical key that subjects indicated for the ten sounds. A global average for all stimuli (out of 120) was computed, as were average scores (out of 30) were for each of the listening environments (*General Home, Kitchen, Office, and Outdoors*). Confusion matrices were generated for each of the listening environments, displaying the prevalence of response confusion and competition. Given the 10 possible response options for each of the listening environments, chance is considered 1 in 10.

Talker gender identification was scored as correct or incorrect, and given the two alternatives (“male” or “female”) chance was considered to be 1 in 2. Similarly, talker discrimination was also scored as correct or incorrect out of two (“same” or “different”).

Results

Pre- and posttest comparisons

Overall, performance during pretest was poor for both groups but improved after training (Figure 1). A two-way repeated measures ANOVA comparing performance across experimental phases for the two training groups revealed a significant main effect of Phase ($F(1,46) = 396.14, p < 0.001$) indicating that across all participants, performance in posttest was significantly higher than in pretest. A significant main effect of Group ($F(1,46) = 7.54, p = 0.009$) was also observed, indicating that subjects who received explicit training on the processed stimuli performed significantly better overall on the posttest compared to those who only transcribed the unprocessed stimuli. A significant interaction was also observed between Group and Phase $F(1,46) = 34.05, p < 0.001$, which suggests that training differentially

affected pre- to posttest gains, and that one group performed significantly better than the other at posttest.

A one-way ANOVA on the pretest scores failed to reveal main effects of Group ($F(1, 46) = 1.86, p = 0.18$) indicating that subjects in both the control group ($M = .37, SD = .13$) and experimental group ($M = .41, SD = .092$) performed equally well during the pretest. In order to ensure that the posttest scores were not overly influenced by pretest scores, one-way ANCOVA was performed to determine the differences between groups. A significant main effect of Group ($F(1, 45) = 33.664, p < 0.001$) was observed even after pretest performance was factored out, indicating that the differences between the groups were driven by the training manipulation, rather than by differences in performance during pretest. Subjects who transcribed vocoded sentences during training ($M = .61, SD = .09$) improved significantly more than subjects in the control group who transcribed unprocessed sentences ($M = .48, SD = .14$). A one-way ANCOVA on the pre-/posttest difference scores also revealed a significant main effect of Group ($F(1, 45) = 33.664, p < 0.001$), indicating that the subjects who received explicit feedback on the vocoded sentences ($\Delta = .20$) showed significantly more improvement at posttest than those who were only exposed to the unprocessed stimuli ($\Delta = .11$) even after pretest performance was factored out.

Gender and talker discrimination

The transfer of perceptual learning to identification of the nonlinguistic aspects of speech (gender identification and talker discrimination) is shown in Figure 2. For gender discrimination, subjects who transcribed the unprocessed sentences during training ($M = .77, SD = .16$) performed as well as subjects who received explicit training on the processed stimuli ($M = .78, SD = .15$). One sample t-tests comparing performance of each group to chance (.50) indicated that both the trained ($t(23) = 10.93, p < .001$) and untrained ($t(23) = 10.94, p < .001$) groups performed significantly better than chance. A 2×2 ANCOVA comparing the effects of training and talker gender on gender identification failed to reveal a significant main effect of group ($F(1, 93) = .673, p = 0.414$), indicating that both groups performed equally well regardless of training. A significant main effect of talker Gender ($F(1, 93) = 15.12, p < .001$) was observed, indicating that subjects were more accurate at identifying the gender of male talkers ($M = .83, SD = .14$) than of female talkers ($M = .71, SD = .14$).

Performance on the talker discrimination task was relatively poor compared to the gender identification task (Figure 2). Subjects who transcribed unprocessed sentences during training ($M = .57, SD = .06$) performed identically to subjects trained on the noise vocoded versions of the sentences ($M = .57, SD = .04$) ($F(1, 46) = .062, p = 0.845$). One-sample t-tests revealed that both the trained ($t(23) = 7.28, p < .001$) and untrained ($t(23) = 5.49, p < .001$) listeners performed significantly better than chance (.50) on the talker discrimination task. Although better than chance, talker discrimination scores were uniformly low. Normal hearing listeners perform near ceiling on the same task using unprocessed stimuli (Cleary, Pisoni & Kirk, 2005). The low overall performance of the participants in the present study, therefore, suggests that the signal processing conditions imposed here may remove some of the characteristics that distinguish speakers of the same sex, making discrimination significantly more difficult than in the unprocessed signal.

Environmental sound identification

Figure 3 provides a summary of the two groups' performance on the environmental sound identification transfer task. Across all stimuli, listeners who received training ($M = .55, SD = .14$) performed better on the environmental sound identification task than listeners who received no training on the spectrally degraded speech ($M = .50, SD = .14$). A two-way ANOVA was conducted using Group (trained or untrained) as the between subjects variable and Place

(Office, Kitchen, Outdoors, and Home) as within subjects variables. A significant main effect of Group was observed ($F(1, 184) = 6.491, p = .012$), indicating that subjects who received training on the noise vocoder performed significantly better than those who were merely exposed to the unprocessed versions of the sentences.

A significant main effect was also observed for Place ($F(3, 184) = 9.271, p < .001$) indicating that subjects performance at identifying environmental stimuli depended partly on which source location they are normally heard in. Post hoc comparisons of the identification scores across the different sound categories (Figure 3) revealed that sounds from the Kitchen were identified least accurately ($M = .45, SD = .13$) followed by sounds from the Home ($M = .49, SD = .11$), Office sounds ($M = .56, SD = .16$) and Outdoor sounds ($M = .58, SD = .14$). Post hoc Bonferroni tests revealed that Outdoor sounds were identified correctly significantly more often than Kitchen or Home sounds ($p < 0.001, p = 0.016$, respectively), but were identified as often as Office sounds ($p = 1.00$). Office sounds were correctly identified significantly more often than Kitchen sounds ($p = 0.016$), but as often as Home sounds ($p = 0.072$). The identification of Kitchen sounds did not differ from the identification of Home sounds ($p = 1.00$).

The interaction between Group and Place approached significance ($F(3, 184) = 2.240, p = .085$), warranting further investigation of the sound categories. Performance on each of the categories was compared across training Groups using a series of one-way ANOVAs. For Outdoor sounds, a significant main effect of Group was observed ($F(1, 46) = 6.951, p = 0.011$), indicating that listeners who received training ($M = .63, SD = .09$) performed significantly better than subjects who received no training ($M = .53, SD = .16$). The sounds from the Outdoor category that showed the largest differences between the trained and untrained groups (differences in accuracy of greater than 10%) were *birds singing* ($d = .19$), *babbling brook* ($d = .14$), *airplane* ($d = .14$), *car starting* ($d = .14$), *dog barking* ($d = .11$), and *fire siren* ($d = .11$). For the remaining sounds, differences between groups were between 8 and 6%, and all but one sound showed positive difference, scores indicating that the trained listeners performed better than the untrained listeners. The exception to this was the *car horn*, which showed a marginal difference favoring the untrained listeners ($d = -.03$).

A significant main effect of Group was also observed for Office sounds ($F(1, 46) = 4.322, p = 0.043$), indicating a significant difference in the performance of the trained ($M = .61, SD = .13$) and untrained ($M = .52, SD = .18$) subjects. The sounds from the Office category that showed the largest differences between the trained and untrained groups were *typing on keyboard* ($d = .36$), *smoke alarm* ($d = .22$), *crowd talking* ($d = .13$), and *fan* ($d = .11$). For the remaining sounds, differences between groups were between 8% and 3%, and all but one sound showed positive difference scores favoring the trained listeners. The exception was *photocopier*, which was identified correctly more often by untrained than trained listeners ($d = -.15$).

No main effects of Group were observed for the Kitchen ($F(1, 46) = .356, p = 0.553$) or Home ($F(1, 46) = .342, p = 0.361$) categories, indicating that training did not affect subjects' performance in either group (Kitchen: $M_T = .46, SD = .13, M_U = .44, SD = .12$; Home: $M_T = .48, SD = .11, M_U = .50, SD = .12$). For the Kitchen sounds, only the *fire alarm* ($d = .15$) and *telephone* ($d = .15$) exceeded 10% differences between groups that favored the trained listeners. All other sounds had difference scores between 7% and -9%, except for *dishes clanging*, which favored the untrained listeners ($d = -.15$). For the Home sounds, only the *dog barking* ($d = .15$) showed difference scores that exceeded 10% and favored the trained listeners. All other sounds were between 3% and -7%, except for *knock on door* ($d = -.12$) and *air conditioner* ($d = -.14$), which favored the untrained listeners.

Examination of the confusion matrices for both the trained and untrained subjects (Appendix A) revealed substantial variability in subjects' abilities to identify sounds across the four categories. Each matrix displays the probabilities of each different response alternative given a specific stimulus. The environmental sounds used in the present study were drawn from the database of Reed and Delhorne (2005), who performed a variety of detailed acoustic measurements on the stimuli. Reed and Delhorne categorized the stimuli according to their temporal properties (overall duration of the signal, presence of one or more transients, and the ratio of the burst duration to the total stimulus duration) and their spectral properties (location of frequency band containing the highest energy) (Reed & Delhorne, 2005). Several interesting patterns of perceptual confusions were made by both groups of listeners in the present study. In the Outdoor sound category, the three most accurately identified sounds for both groups of listeners were the *helicopter*, *dog bark*, and *car starting*, all of which have distinct temporal components (periodic patterns for the helicopter rotors, characteristic transients for the car motor, and repeating burst patterns for the dog barking). On the other hand, both groups of listeners identified *siren* as *plane*, and *thunder* as *babbling brook*. These sounds lack transients and contain burst durations equal to the total stimulus duration, but contain complex harmonic spectra (as in the case of the siren), a higher maximum frequency (as in the case of the jet engine), or complex spectral components (as in the case of the rumbling thunder and the rolling water of the babbling brook), which reliably distinguish the sounds. In the Office sound category, both groups of listeners frequently reported hearing the *water cooler* sound instead of *fan*, *copy machine*, and *paper rustling*. The two most accurately identified sounds for both the trained and untrained listeners were the *file door slam* and *footsteps*, which were both classified as transient sounds and have distinct temporal components (a pronounced impact of the file drawer, and the repeated regular gait of the foot falls).

From the examination of the confusion data, it is apparent that subjects were having more difficulty identifying environmental sounds that required the identification of fine spectral information, and less difficulty on the temporally distinct sounds (according to the acoustic analyses conducted by Reed and Delhorne, 2005). To examine the effect of acoustic features on environmental sound identification, all environmental sounds were classified according to their temporal properties (overall signal duration, transiency, and ratio of burst to total duration), spectral properties (frequency region with peak intensity) and intensity properties (RMS amplitude) according to the methods of Reed and Delhorne (2005). Of these five acoustic factors, Transiency ($r = .405, p < 0.001$) and Amplitude ($r = .274, p = 0.014$) were positively correlated with correct environmental sound identification, indicating that sounds that were more transient in nature or higher in amplitude were identified correctly significantly more often than less transient, lower amplitude sounds (Table 2). A significant negative correlation was observed between Burst ratio (the duration of the burst divided by the overall duration of the stimulus) and correct environmental sound identification ($r = -.424, p < 0.001$) indicating that sounds that contained bursts that were short in comparison to the duration of the overall signal were identified significantly more often than sounds that contained longer bursts. This may be confounded with transiency, however, since multiple bursts will necessarily impart different temporal characteristics on the signal, such as sounds that only have a single burst that is equal to the length of the stimulus. The significant correlation between transiency and burst duration underscores this possibility. Total signal duration ($r = -.147, p = .193$) and frequency region at peak intensity ($r = -.035, p = 0.760$) were not correlated with environmental sound recognition accuracy, indicating that these characteristics may not influence recognition.

Performance on the environmental sound identification was compared with performance on transcription accuracy at pre- and posttest using a linear regression analysis. In order to account for differences in posttest gains across subjects, scores from both phases were entered as independent variables in the regression model. Across all subjects, performance at pre- and posttest was significantly correlated with environmental sound recognition ($\beta = .537, F(2,$

45) = 9.135, $p < 0.001$). Factor analysis revealed that pretest performance accounted for very little of the variance in the model ($\beta = .120$, $t = .491$, $p = 0.626$) whereas performance at posttest accounted for most of the variance in environmental sound identification ($\beta = .637$, $t = 2.601$, $p = 0.013$). The curve fit for the posttest data (Figure 4) reveals a linear relationship between posttest transcription scores and environmental sound identification, with subjects who perform well on one, performing well on the other. This relationship suggests that there may be substantial overlap in the information required for successful speech transcription and environmental sound identification.

In order to investigate further the effects of training on environmental sound identification, pre-/posttest difference scores were also compared with environmental sound identification using a linear regression model. Difference scores were also significantly correlated with performance on environmental sound identification ($\beta = .352$, $t = 2.553$, $p = 0.014$), indicating that participants who showed the most pre- to posttest improvement performed better on the environmental sound identification task (Figure 5). Two outliers who showed abnormally low performance on the environmental sound identification test but relatively normal pre-/posttest gains can be noted in the scatter plot. When these two subjects are excluded from the regression model, the strength of the relationship between environmental sound identification and gains from training increases ($\beta = .492$, $t = 14.808$, $p < 0.001$). Since the relationship between pre-/posttest gains and environmental sound identification was significant before and after removing these outliers, we report the full dataset here. These findings indicate that the correlations observed between the posttest performance and environmental sound identification is not simply due to a few subjects who were performing well on environmental sound identification, but rather were driven by the amount of pre- to posttest improvement. This may be a critically important finding for rehabilitative protocols for cochlear implant users, because it indicates that general auditory abilities are linked with fundamental speech perception abilities and suggests that training on one may lead to transfer to and/or enhancement of the other.

Discussion

Several findings on auditory perceptual learning and the perception of speech and nonspeech stimuli were uncovered in the present study. First, although both training groups displayed significantly higher performance at posttest due to general practice effects, subjects who were explicitly trained with the noise vocoded stimuli performed significantly better than the control group. This suggests that explicit training with vocoded speech produces effects that are above and beyond those expected from generalized task familiarity or global practice effects. One possible limitation of the present study that could temper some of the conclusions made arises from the inclusion of a control group that was not exposed to the degraded materials during training. Previous research has demonstrated that adaptation to noise vocoded speech occurs rapidly (Davis et al., 2005; Loebach and Pisoni, 2008), and that it is not merely the effect of passive exposure to the vocoded materials or simple practice effects that drive such improvement (Loebach, Pisoni, & Svirsky, In Press) but the tasks actively employed during training (Loebach, Bent & Pisoni, 2008). Therefore, in light of the findings from past work, it is unlikely that exposure effects alone could have accounted for the differences in perceptual learning across group observed in the present study. Moreover, recent work (Loebach, Pisoni, & Svirsky, In Press) has directly compared exposure effects to training effects and found similar pre- to posttest improvement for participants trained with vocoded auditory and orthographic feedback compared to participants exposed to the same vocoded materials without feedback as was found in the present study. Therefore, the absence of a true control condition in the present study does not adversely affect the interpretations that can be made from the data.

Second, subjects were quite accurate (77% correct) at identifying talker gender and no differences in performance were observed between subjects who were explicitly trained with

the vocoded stimuli and those who merely heard the unprocessed stimuli during training. The finding that subjects did not require explicit training to perform above chance in the gender identification task suggests that they may have been utilizing similar processes to identify the gender of vocoded voices as they would to identify the gender of unprocessed voices. The lack of generalization from speech training and the finding that subject performance was not near ceiling, however, suggests that specific training on gender identification may be required to achieve high levels of gender recognition accuracy. Previous studies have demonstrated that listeners need only short samples of speech materials (isolated words, vowels or fricatives) to accurately identify talker gender (Lass, Hughes, Bowyer, Waters, & Bourne, 1976; Schwartz, 1968). Given the long samples of speech provided by the meaningful Harvard sentences, it is not surprising that subject accuracy on gender identification would exceed chance even though the stimuli were highly degraded. Finally, past research on gender identification with vocoded materials demonstrated that subject performance was uniformly high and comparable with that of cochlear implant users (Fu, Chinchilla and Galvin, 2004). Since the results of the present study are comparable to those observed by Fu and colleagues (2004), the data should also extend to cochlear implant users as well even though such individuals were not tested in the present study.

Third, while slightly above chance, performance on talker discrimination was poor overall, and did not differ across groups, suggesting that specific training may be necessary to learn to reliably discriminate talkers by voice. Given that all pairs of talkers were of the same gender, subject difficulty in reliably discriminating talkers suggests that the acoustic information that specifies unique voices may not be temporal in nature and instead may rely on fine detailed spectral cues that are not preserved in vocoder simulations of cochlear implants. Previous studies have demonstrated that talker discrimination in pediatric CI users is considerably more difficult when each talker produces a different sentence than when they produce the same sentence (Cleary & Pisoni, 2002; McDonald *et al.*, 2003). It is possible that the poor performance of the subjects in the present study was a consequence of such variability since all talker pairs consisted of different utterances. Although performance would have conceivably been better if the same sentences were used for both talkers in a pair, this method was chosen because it is presumed to be more clinically applicable and externally valid than presenting linguistically identical sentences to listeners in a talker discrimination task.

Additionally, although talker discrimination was low in the present study, it was significantly greater than chance, and comparable to cochlear implant users from previous studies, who averaged 57% correct talker discrimination in the variable sentence condition (Cleary and Pisoni, 2002). Although previous research has demonstrated that CI users perform poorly on talker identification (25% correct, Vongphoe & Zeng, 2005), differences in task difficulty could limit extension to the present study. Vongphoe and Zeng (2005) used a talker identification paradigm, where participants were asked to identify a talker by voice. This task requires that participants while the present study used a simpler talker discrimination paradigm. Past research comparing talker identification training using vocoded sentences, however, revealed similar levels of performance (55% correct after 2 days of training) to those demonstrated for talker discrimination in the present study (Loebach, Bent & Pisoni, 2008). Therefore, explicit training may be required to produce robust performance in more difficult tasks, such as talker identification by voice (Loebach, Bent & Pisoni, 2008). An additional possibility is that the information required for accurate talker identification may not be well represented by vocoder models of cochlear implant speech processors, or by cochlear implant speech processors themselves. This possibility requires further empirical research before firm conclusions can be made.

Compared to noise vocoded speech, talker identification and discrimination with sinewave speech is far more accurate (Remez, Fellowes & Rubin, 1997). Several critical differences

exist, however, between these two synthesis methods. Sinewave speech mimics formant movement, preserving overall changes in frequency and amplitude in the dynamically changing narrow band sinusoids at the expense of temporal envelope information (Remez *et al.*, 1981). Noise vocoded speech effectively sums frequency information over larger portions of the spectrum into broad bands removing much information about formant trajectory (Teoh *et al.*, 2003). In some cases, multiple harmonics or even multiple formants may be represented in a single noise band, thus removing critically important spectral cues such for talker identity, such as formant spacing (Klatt & Klatt, 1990). Recent work using sinewave vocoders (Gonzalez & Oliver, 2005), has demonstrated that talker and gender discrimination was significantly better for vocoders using sinewave carriers when compared to noise band carriers. It remains possible that training with sinewave rather than noise vocoders may foster transfer of perceptual learning to talker discrimination tasks, but further research will be necessary to determine the effect of signal carrier.

Finally, irrespective of training, all participants were fairly accurate in identifying spectrally degraded environmental sounds. Subjects who received explicit training with vocoded sentences, however, performed significantly better than untrained control subjects, suggesting that auditory perceptual learning of speech transfers to nonspeech environmental sounds. Sounds with distinct transient or temporal properties appeared to be easiest for listeners to identify in both training groups. The significant correlations between percent correct recognition and the presence of temporal features such as transiency and the ratio of the burst to total signal duration provide further support for this claim, suggesting that subjects were utilizing the residual temporal information in the signal in the environmental sound identification task. These results are consistent with earlier findings from adult cochlear implant users (Reed & Delhorne, 2005) and normal-hearing listeners trained to identify vocoded versions of environmental sounds (Gygi *et al.*, 2004). Moreover, the significant correlations between posttest performance and environmental sound identification scores indicate that subjects who were better able to utilize the residual acoustic information to identify environmental sounds were also more accurate at transcribing the vocoded sentences. The significant correlation between the pre-/posttest difference scores and environmental sound identification further demonstrates that training on speech transfers to the perception of environmental sounds (a nonspeech task). Taken together, the present results suggest that the ability to utilize general acoustic information supports both speech and nonspeech sound identification for vocoded stimuli.

The present results suggest that the ability to transfer perceptual learning from a speech transcription task to other nonlinguistic tasks may be partially task dependent. Although training did transfer to environmental sound identification, it was limited to only two sound categories: Outdoor sounds and Office sounds. In the Home and Kitchen sound categories, the untrained listeners performed as well as the trained listeners. The trained listeners were better able to utilize cues in the temporal envelope for the identification of environmental sounds from the Office and Outdoor categories as a consequence of training. Reed and Delhorne (2005) found that while both high and low performing CI users performed equally well on transient sounds, high performing users were significantly better at identifying sounds without distinct temporal cues than low performing users. The training effects observed in the present study suggests that explicit training with speech generalizes to at least some categories of environmental sounds. To some extent, the converse is also true, and as Loebach and Pisoni (2008) have recently demonstrated, training with environmental sounds also generalizes to the perception of speech, but not the reverse. A possible explanation for this inconsistency lies in the differences in methodology used to test generalization to environmental sounds. In the previous work, participants were asked to identify environmental sounds in an open set paradigm, where they were presented with a sound and asked to generate a spontaneous description of the source (Loebach & Pisoni, 2008). In the present study, a closed set forced

choice paradigm was used that is a direct replication of Reed and Delhorne (2005). The ability to generate a label for a sound source may draw on different cognitive systems than those used to select the most appropriate label from a list; therefore, the results of the two studies are not directly comparable. Finally, the two studies had very different aims. The goal of the previous study was to test the effects of the bidirectionality of training with speech and environmental sounds (Loebach & Pisoni, 2008), while the goal of the present study was to examine how training affects the recognition of environmental based on the acoustic elements that are enhanced through training with speech. Therefore, further research will be required to determine the extent of transfer of perceptual learning to and from environmental sounds to speech, what type of tasks will be most efficient and most representative of real life listening situations (open vs. closed set response alternatives), and whether such training would be beneficial for cochlear implant users.

The results of the present study are of both theoretical and clinical relevance. The transfer of perceptual learning from a speech transcription task to an environmental sound identification task is theoretically important because it suggests that the perceptual learning induced during training was not specific to speech and implicates the involvement of a more domain-general process. This result complements previous findings, which suggest that the phonetic and indexical information carried in the speech stream interact during speech processing (Nygaard *et al.*, 1994; Nygaard & Pisoni, 1998; Remez *et al.*, 1997). The results of the present study are also clinically important since the ability to identify and recognize environmental sounds has potential real-world significance for CI users outside the clinic and laboratory. The benefit of training applied to both speech and some but not all nonspeech tasks, suggesting that rehabilitation paradigms for newly implanted individuals should not only focus on speech, but should include a wide variety of training tasks and acoustic materials.

As with all studies that use cochlear implant simulations, great care should be taken to ensure that the results obtained here apply to CI users as well. Although the training period was in the present study was very brief (compared to cochlear implant users, who have many hours of experience using their devices), the results highlight the malleability of perceptual learning and the potential benefits of training for newly implant individuals. A future direction for this research should include investigations of how gender identification, talker discrimination and environmental sound recognition skills can be improved in CI users. Presently, it is unclear whether task-specific training would be required to improve talker discrimination or if training on other skills such as frequency or electrode discrimination would enhance listeners' abilities to discriminate the identity of talkers. In Reed and Delhorne's original study (2005), cochlear implant users could be divided into low performing (65% correct) and high performing (90% correct) groups based on their scores on the environmental sounds test. At 55% correct, the trained listeners in the present study appear to perform more like the low performing cochlear implant users from Reed and Delhorne (2005). Moreover, both groups of CI users performed best on the Office and Outdoor categories (Reed & Delhorne, 2005), similar to the normal hearing listeners in the present study. The finding that normal hearing listeners can achieve similar performance to cochlear implant users in a brief training session (compared to the low performing CI users who had an average of 6 years of experience with their implants) further underscores the importance of training for new CI users. Future studies on this topic will be necessary to elucidate the specific rehabilitation paradigms that should be employed with CI users. In addition, future research will need to determine the extent of bidirectionality of training. The present study demonstrated that training on speech perception transfers to some types of nonspeech stimuli. Recently, Loebach and Pisoni (2008) found that training on nonspeech stimuli also transfers to speech perception. Further research will need to determine whether training on talker or environmental sound identification in CI users will also transfer to speech perception.

Conclusions

In summary, the results of the present study suggest that training normal-hearing listeners to identify speech processed with an acoustic simulation of a cochlear implant transfers to the closed set identification of environmental sounds. Although both groups performed equally well at pre-test, participants trained with spectrally degraded sentences showed significantly higher performance at posttest than those participants trained with undegraded materials. Moreover, although all subjects performed equally well on environmental sounds that carried distinct temporal cues, subjects in the training group performed significantly better on stimuli that did not have a distinct temporal component, presumably due to training. In contrast, training did not transfer to the gender identification task presumably because the information necessary for such a task is not well preserved by the vocoder, and explicit training may be required to provide high levels of performance on the gender identification task. Perceptual learning did not transfer to the talker discrimination task, and although both groups performed above chance, their performance was uniformly low. It is possible that the important information about talker identity, such as formant spacing, was not well preserved by the vocoder. The findings of the present study also suggest that the perceptual learning provided by speech based training paradigms for cochlear implant users could be transferred to environmental sound identification, but other tasks such as talker discrimination may require explicit task specific training to elicit improvement. Future studies should be conducted to determine what training methods are most effective and to determine whether signal-processing methods could be modified to provide cochlear implant users with more detailed frequency information to use in identifying talkers as well as environmental sounds.

Acknowledgments

This research was supported by NIH-NIDCD research grants DC00111 and DC03937, training grant DC00012, and the American Hearing Research Foundation. We are grateful for the time and support from Speech Research Lab members who participated in the piloting of these experiments. We also thank Shivank Sinha and Luis Hernandez for their technical assistance on this project. We would like to thank Rose Burkholder-Juhasz for her assistance in data collection and for her contributions to an earlier draft of this manuscript.

Appendix A

I

Stimulus-response confusion matrices for the Outdoor stimulus set for trained and (untrained) listeners. Cell contents represent the probabilities of a response given a specific stimulus. Stimuli on the diagonal indicate instances where the stimulus and response were the same, indicating a correct response.

Response										
	Birds	Brook	Car	Chopper	Dog	Horn	Keys	Plane	Siren	Thunder
Stimulus										
Birds	.72 (.53)	-- (.03)	.06 (.08)	.08 (.12)	.01 (.03)	.03 (.01)	.08 (.08)	.01 (.01)	-- (.03)	-- (.04)
Brook	.03 (.06)	.68 (.54)	.04 --	.17 (.17)	-- --	-- --	.04 (.03)	.03 (.06)	.01 (.07)	-- (.04)
Car	-- --	.01 (.10)	.92 (.78)	.01 --	-- --	.03 --	-- --	.01 (.01)	-- --	.01 (.04)
Chopper	-- --	-- --	.01 (.10)	.99 (.93)	-- --	-- --	-- --	-- --	-- --	-- --
Dog	-- --	-- (.01)	-- (.01)	.03 (.06)	.90 (.79)	.01 (.04)	.01 (.01)	-- --	-- (.03)	.04 --
Horn	-- --	.03 (.10)	.10 (.03)	.01 --	.19 (.14)	.40 (.43)	.01 (.03)	.11 (.07)	.04 (.01)	.10 (.14)
Keys	.07 (.17)	.07 (.04)	.03 (.03)	.03 (.06)	-- --	.03 --	.69 (.64)	.01 (.01)	.03 --	.01 --
Plane	-- (.04)	.15 (.22)	.19 (.17)	.11 (.08)	-- --	.03 --	.03 (.01)	.25 (.11)	.07 (.11)	.17 (.22)
Siren	.06 (.03)	.10 (.04)	.01 (.06)	.12 (.19)	.01 --	-- --	-- --	.35 (.35)	.32 (.21)	.03 (.08)
Thunder	.01 --	.40 (.44)	.07 (.04)	-- --	-- (.06)	.01 (.01)	.07 (.04)	.03 (.03)	-- (.01)	.40 (.32)

II

Stimulus-response confusion matrices for the Office stimulus set for trained and (untrained) listeners. Cell contents represent the probabilities of a response given a specific stimulus. Stimuli on the diagonal indicate instances where the stimulus and response were the same, indicating a correct response.

Response										
	Cooler	Copier	Crowd	Fan	FileDoor	Papers	Telephone	Smoke	Steps	Typing
Stimulus										
Cooler	.64 (.61)	.06 (.01)	.01 (.03)	.14 (.15)	-- --	.04 (.03)	.10 (.06)	.01 (.03)	-- --	-- (.04)
Copier	.32 (.14)	.15 (.31)	-- --	.12 (.10)	.07 (.06)	.21 (.25)	.03 (.03)	.01 (.03)	.01 (.03)	.06 (.01)
Crowd	.13 (.14)	.01 (.03)	.79 (.67)	.01 (.10)	-- --	.06 (.04)	-- --	-- --	-- --	-- --
Fan	.33 (.31)	.11 (.14)	-- (.01)	.25 (.14)	-- --	.18 (.19)	.04 (.10)	.06 (.01)	.01 (.01)	-- (.01)
FileDoor	-- --	.04 (.06)	.01 --	.01 (.01)	.81 (.78)	.04 (.08)	.03 --	-- --	.01 --	.03 (.01)
Papers	.11 (.36)	.22 (.04)	.03 (.07)	.04 --	.04 (.04)	.42 (.33)	.01 (.04)	-- --	-- --	.10 (.10)
Telephone	.04 (.08)	.06 (.11)	-- (.01)	.31 (.19)	-- --	.04 (.01)	.47 (.42)	.07 (.07)	.01 --	-- (.04)
Smoke	-- --	.01 (.12)	-- --	.08 (.10)	-- --	-- --	.03 (.08)	.86 (.64)	-- (.01)	-- --
Steps	-- (.06)	.01 (.01)	-- --	.01 (.01)	-- --	-- --	.01 (.01)>	-- --	.89 (.81)	.07 (.06)
Typing	.01 (.06)	.03 (.08)	-- (.01)	.04 (.22)	-- --	.083 (.08)	-- (.04)	-- --	-- --	.83 (.47)

III

Stimulus-response confusion matrices for the Home stimulus set for trained and (untrained) listeners. Cell contents represent the probabilities of a response given a specific stimulus. Stimuli on the diagonal indicate instances where the stimulus and response were the same, indicating a correct response.

Response										
	Air	Baby	Dog	Door	Flush	Knock	Music	Rain	Vacuum	Voice
Stimulus										
Air	.08 (.22)	.01 --	-- --	-- --	.12 (.15)	-- --	.04 (.03)	.65 (.50)	.07 (.07)	-- --
Baby	.06 (.01)	.26 (.33)	.06 (.07)	.01 --	.08 (.1)	.03 --	.12 (.1)	.03 --	.24 (.18)	.1 (.18)
Dog	-- (.01)	-- --	.92 (.76)	.01 (.08)	-- (.01)	.03 (.08)	-- (.14)	-- --	.01 --	.03 (.01)
Door	-- --	-- --	.01 --	.79 (.92)	.014 --	.15 (.08)	.01 --	.01 --	-- --	-- --
Flush	.06 (.03)	.03 (.07)	-- --	.1 (.04)	.46 (.49)	-- --	.04 (.08)	.26 (.21)	.04 (.06)	.01 (.01)
Knock	.11 (.08)	-- --	-- --	-- (.03)	.06 (.03)	.72 (.71)	-- (.06)	.07 (.08)	.01 --	.01 (.01)
Music	.03 (.07)	.03 (.01)	.08 (.06)	-- --	.39 (.29)	.04 (.06)	.19 (.17)	.06 (.14)	.17 (.08)	.01 (.13)
Rain	.11 (.04)	-- --	-- --	-- --	.08 (.06)	-- --	-- (.03)	.76 (.82)	.04 (.04)	-- --
Vacuum	.17 (.06)	.03 (.04)	-- --	.01 --	.12 (.11)	-- --	.06 (.07)	.35 (.5)	.25 (.22)	.01 --
Voice	.1 (.17)	.19 (.11)	.03 (.03)	.04 (.04)	.01 (.06)	.01 (.01)	.18 (.1)	.01 (.01)	.01 (.04)	.40 (.40)

IV

Stimulus-response confusion matrices for the Kitchen stimulus set for trained and (untrained) listeners. Cell contents represent the probabilities of a response given a specific stimulus. Stimuli on the diagonal indicate instances where the stimulus and response were the same, indicating a correct response.

Response										
	Smoke	Doorbell	Cat	Cupboard	Dishes	Dishwasher	Telephone	Steps	Timer	Water
Stimulus										
Smoke	.74 (.58)	.07 --	.03 (.01)	-- (.01)	.01 (.04)	.03 (.08)	.03 (.01)	-- (.01)	.10 (.17)	-- (.07)
Doorbell	.03 (.06)	.15 (.15)	.08 (.06)	.10 (.15)	.12 (.06)	.10 (.25)	.10 (.03)	-- (.01)	.07 (.07)	.24 (.12)
Cat	.03 --	.06 (.07)	.15 (.17)	.12 (.14)	.14 (.07)	.17 (.12)	-- (.01)	.04 (.03)	.03 (.03)	.26 (.35)
Cupboard	-- --	.01 (.03)	.01 --	.82 (.81)	.10 (.04)	-- --	.01 (.03)	-- (.03)	.04 (.06)	-- (.01)
Dishes	-- --	.01 (.03)	.01 (.01)	.08 (.04)	.58 (.74)	.07 (.07)	.03 (.03)	.14 (.04)	.04 (.03)	-- --
Dishwasher	.03 --	.01 (.03)	.01 (.01)	.01 (.01)	.01 --	.19 (.14)	.01 (.03)	.03 --	-- (.03)	.67 (.74)
Telephone	.03 (.15)	.03 (.03)	.07 (.04)	.01 (.01)	.01 --	.19 (.11)	.49 (.33)	.01 (.01)	.06 (.11)	.10 (.19)
Steps	-- --	-- (.01)	.01 (.03)	-- --	.01 (.03)	.03 --	-- (.03)	.89 (.81)	-- (.01)	.06 (.04)
Timer	.01 (.01)	.04 (.01)	-- --	-- (.01)	.01 (.03)	.11 (.12)	.08 (.04)	-- --	.07 (.03)	.65 (.72)
Water	.01 (.07)	.07 (.01)	.03 --	.01 --	.01 (.01)	.19 (.12)	.06 (.03)	-- --	.06 (.10)	.56 (.65)

REFERENCES

- Ballas JA. Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception and Performance* 1993;19:250–267. [PubMed: 8473838]
- Benson NJ, Lovett MW, Kroeber CL. Training and transfer-of-learning effects in disabled and normal readers: Evidence of specific deficits. *Journal of Experimental Child Psychology* 1997;64(3):343–366. [PubMed: 9073377]
- Bode DL, Oyer HJ. Auditory training and speech discrimination. *Journal of Speech and Hearing Research* 1970;13(4):839–855. [PubMed: 5491358]
- Bradlow AR, Torretta GM, Pisoni DB. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication* 1996;20:255–272.
- Burkholder, RA. Perceptual Learning with Acoustic Simulations of Cochlear Implants. 2005. Unpublished doctoral dissertation
- Cleary M, Pisoni DB. Talker discrimination by prelingually deaf children with cochlear implants: preliminary results. *Annals of Otology, Rhinology, and Laryngology* 2002;189:113–118.
- Cleary M, Pisoni DB, Kirk KI. Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. *Journal of Speech, Language, and Hearing Research* 2005;48(1):204–223.

- Davis, M.; Taylor, K.; Johnsrude, I.; Carlyon, B. How do cochlear implant users learn to understand speech? Transfer of learning between different carriers with vocoded speech; Poster presented at the British Society of Audiology; London, England. Sep. 2004
- Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C. Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General* 2005;134(2):222–241. [PubMed: 15869347]
- Deguchi T, Kagami R, Hiki S. Pitch perception by hearing-impaired children: Possibility of improvement through auditory training. *Japanese Journal of Special Education* 1981;18(4):70–78.
- Delay E. Cross-modal transfer effects on visual discrimination depends on lesion location in the rat visual system. *Physiology & Behavior* 2001;73(4):609–620. [PubMed: 11495666]
- Delhommeau K, Micheyl C, Jouvent R, Collet L. Transfer of learning across durations and ears in auditory frequency discrimination. *Perception and Psychophysics* 2002;64(3):426–436. [PubMed: 12049283]
- Dorman MF, Loizou PC, Rainey D. Speech intelligibility as a function of the number of channels of stimulation for signal processing using sine-wave and noise bands. *Journal of the Acoustical Society of America* 1997;102(4):2403–2411. [PubMed: 9348698]
- Fenn KM, Nusbaum HC, Margoliash D. Consolidation during sleep of perceptual learning of spoken language. *Nature* 2003;425:614–616. [PubMed: 14534586]
- Fu Q-J, Chinchilla S, Galvin JJ. The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology* 2004;5:253–260. [PubMed: 15492884]
- Fu Q-J, Galvin JJ. The effects of short-term training for spectrally mismatched noise-band speech. *Journal of the Acoustical Society of America* 2003;113(2):1065–1072. [PubMed: 12597199]
- Fu Q-J, Shannon RV. Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *Journal of the Acoustical Society of America* 1999;105:1889–1900. [PubMed: 10089611]
- Gfeller K, Mehr M, Witt S. Aural rehabilitation of music perception and enjoyment of adult cochlear implant users. *Journal of the Academy of Rehabilitative Audiology* 2001;34:17–27.
- Gonzalez J, Oliver JC. Gender and talker identification as a function of the number of channels in spectrally reduced speech. *Journal of the Acoustical Society of America* 2005;118(1):461–470. [PubMed: 16119365]
- Greenspan SL, Nusbaum HC, Pisoni DB. Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1988;14(3):421–433.
- Gygi B, Kidd G, Watson C. Spectral-temporal factors in the identification of environmental sounds. *Journal of the Acoustical Society of America* 2004;115(3):1252–1265. [PubMed: 15058346]
- Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 1988;75:800–802.
- Hunstad E. Visual reading and cross-modal transfer of learning in congenitally blind humans with residual light projection. *Scandinavia Journal of Educational Research* 1985;29(1):17–41.
- IEEE. IEEE Recommended Practice for Speech Quality Measurements. Institute of Electrical and Electronic Engineers; New York: 1969.
- Kaiser, AR.; Svirsky, MA. Using a personal computer to perform real-time signal processing in cochlear implant research; Paper presented at the Proceedings of the IXth IEEE-DSP Workshop; Hunt, TX. Oct 15-18. 2000
- Karl J, Pisoni D. The role of talker-specific information in memory for spoken sentences. *Journal of the Acoustical Society of America* 1994;95:2873.
- Kennedy DK, Weener P. Visual and auditory training with the cloze procedure to improve reading and listening comprehension. *Reading Research Quarterly* 1973;8(4):524–541.
- Kirk, KI.; Houston, DM.; Pisoni, DB.; Springer, A.; Kim-Lee, Y. Talker discrimination and spoken word recognition by adults with cochlear implants. Poster presented at the 25th Midwinter Meeting of the Association for Research in Otolaryngology; St. Petersburg, Florida. 2002.
- Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 1990;87:820–857. [PubMed: 2137837]

- Kong Y-Y, Cruz R, Jones JA, Zeng F-G. Music perception with temporal cues in acoustic and electric hearing. *Ear and Hearing* 2004;25(2):173–185. [PubMed: 15064662]
- Lass N, Hughes K, Bowyer M, Waters L, Bourne V. Talker sex identification from voiced, whispered, and filtered isolated vowels. *Journal of the Acoustical Society of America* 1976;59:675–678. [PubMed: 1254794]
- Loebach JL, Bent T, Pisoni DB. Multiple Routes to the perceptual learning of speech. *Journal of the Acoustical Society of America* 2008;124(1):552–561. [PubMed: 18646998]
- Loebach JL, Pisoni DB. Perceptual learning of spectrally degraded speech and environmental sounds. *Journal of the Acoustical Society of America* 2008;123(2):1126–1139. [PubMed: 18247913]
- Loebach JL, Pisoni DB, Svirsky MA. Effects of semantic context and feedback on perceptual learning of speech processed through an acoustic simulation of a cochlear implant. *Journal of Experimental Psychology: Human Perception and Performance*. (In Press).
- McDonald, CJ.; Kirk, KI.; Krueger, T.; Houston, D.; Sprunger, A. Talker discrimination by adults with cochlear implants. Poster presented at the 26th Midwinter Meeting of the Association for Research in Otolaryngology; Daytona Beach, Florida. 2003.
- Moore, BCJ. *An Introduction to the Psychology of Hearing*. Academic Press; San Diego, CA: 1997.
- Muramoto T. Types of comprehension processes and the preliminary knowledge-use effect: Analogical transfer of learning from text. *Japanese Journal of Psychology* 2001;5:429–434. [PubMed: 11883331]
- Murray J. Effects of whole versus part method of training on transfer of learning. *Perceptual and Motor Skills* 1981;53(3):883–889.
- Nakagawa E. Transfer of learning between matching (or non-matching)- to sample and same-different discrimination in rats. *Psychological Record* 2000;50(4):771–805.
- Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker contingent process. *Psychological Science* 1994;5(1):42–45.
- Nygaard LC, Pisoni DB. Talker-specific learning in speech perception. *Perception & Psychophysics* 1998;60(3):355–376. [PubMed: 9599989]
- Pisoni DB, Manous LM, Dedina MJ. Comprehension of natural and synthetic speech: effects of predictability on the verification of sentences controlled for intelligibility. *Computer Speech and Language* 1987;2:303–320.
- Reed CM, Delhorne LA. The reception of environmental-sounds through wearable tactual aids. *Ear and Hearing* 2003;24:528–538. [PubMed: 14663352]
- Reed CM, Delhorne LA. Reception of environmental sounds through cochlear implants. *Ear and Hearing* 2005;26(1):48–61. [PubMed: 15692304]
- Remez RE, Fellowes JM, Rubin PE. Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance* 1997;23(3):651–666. [PubMed: 9180039]
- Remez RE, Rubin PE, Pisoni DB, Carrell TD. Speech perception without traditional speech cues. *Science* 1981;212:947–950. [PubMed: 7233191]
- Robbins, AM. Lesson plan for Lilly. In: Estabrooks, W., editor. *Cochlear Implants for Kids*. A.G. Bell Association; Washington DC: 1998. p. 153-174.
- Rosen S, Faulkner A, Wilkinson L. Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *Journal of the Acoustical Society of America* 1999;106(6):3629–3636. [PubMed: 10615701]
- Roth DA-E, Kishon-Rabin L, Hildesheimer M, Karni A. A latent consolidation phase in auditory identification learning: time in the awake state is sufficient. *Learning & Memory* 2005;12:159–164. [PubMed: 15805314]
- Schwab EC, Nusbaum HC, Pisoni DB. Some effects of training on the perception of degraded speech. *Human Factors* 1985;27(4):395–408. [PubMed: 2936671]
- Schwartz M. Identification of talker sex from isolated, voiceless fricatives. *Journal of the Acoustical Society of America* 1968;43:1178. [PubMed: 5648107]
- Shafiro V. Development of a large-item environmental sound test and the effects of short-term training with spectrally-degraded stimuli. *Ear & Hearing* 2008;29(5):775–790. [PubMed: 18596641]

- Shannon RV, Zeng F, Kamath V, Wygonski J, Ekelid M. Speech recognition with primary temporal cues. *Science* 1995;270(13):303–304. [PubMed: 7569981]
- Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 2002;416:87–90. [PubMed: 11882898]
- Stakhovskaya O, Sridhar D, Bonham BH, Leake PA. Frequency map for the human cochlear spiral ganglion: implications for cochlear implants. *Journal of the Association for Research in Otolaryngology* 2007;8(2):220–233. [PubMed: 17318276]
- Teixeira LA. Timing and force components in bilateral transfer of learning. *Brain & Cognition* 2000;44(3):455–469. [PubMed: 11104537]
- Teoh S-W, Neuberger HS, Svirsky MA. Acoustic and electrical pattern analysis of consonant perceptual cues used by cochlear implant users. *Audiology and Neurotology* 2003;8(5):269–85. [PubMed: 12904682]
- Teoh S-W, Pisoni DB, Miyamoto RT. Cochlear implantation in adults with prelingual deafness. Part II. Underlying constraints that affect audiological outcomes. *Laryngoscope* 2004;114(10):1714–1719. [PubMed: 15454759]
- Watanabe S. Interocular transfer of learning in the pigeon: Visual-motor integration and separation of discriminanda and manipulanda. *Behavioral Brain Research* 1986;19(3):227–232.

Figure 1.

Percent keyword correct transcription scores at pre- and posttest for untrained and trained listeners. Error bars represent the standard error of the mean. The change in percentage points from pre- to posttest for each group is noted with a colored triangle. Asterisks immediately above the bars at posttest indicate significant differences between groups (** $p < 0.01$).

Figure 2. Percent correct gender identification and talker discrimination scores for untrained and trained listeners. Error bars represent the standard error of the mean. The dashed line indicates chance performance (1 in 2, or 50%). All values were significantly higher than chance performance.

Figure 3.

Percent correct environmental sound identification scores for untrained and trained listeners across all stimuli, and for each sound category. Error bars represent the standard error of the mean. Asterisks indicate significant differences between groups (* $p < 0.05$).

Figure 4.

Linear regression of the percent correct keyword transcription during posttest (x-axis) on percent correct environmental sound identification (y-axis). The equation (inset) indicates that subjects' performance on the posttest accurately predicts performance on environmental sound identification after training.

Figure 5.

Linear regression of the difference scores between pre- and posttest (x-axis) on percent correct environmental sound identification (y-axis). The equation (inset) indicates that the amount of improvement that participants showed between pre- and posttest was significantly correlated with performance on environmental sound identification.

Table 1

List of the sound categories and sounds used in the environmental sound identification task. Sounds classified as having transient or distinct temporal properties appear in italics (Reed & Delhorne, 2005)

General Home	Kitchen	Office	Outside
Air conditioner	Smoke alarm	Paper rustling	Birds singing
Baby crying	Cat meowing	<i>File-door slamming</i>	Keys jangling
Dog barking	<i>Cupboard-door slamming</i>	<i>Typing on keyboard</i>	Dog barking
Male voice	Dishes clanging	Telephone	Helicopter
Music	Dishwasher	Photocopier	Airplane
Toilet flushing	Doorbell	Water cooler	Fire siren
Rain	<i>Footsteps</i>	Smoke alarm	<i>Car horn</i>
<i>Knock on door</i>	Oven timer	Crowd talking	Thunder
Vacuum cleaner	Telephone	<i>Footsteps</i>	Babbling brook
<i>Door closing</i>	Running water	Fan	Car starting

Table 2

Correlations between the percent correct recognition scores for the environmental sounds (last column) and the five acoustic features of the stimuli defined by Reed and Delhorne (2005): stimulus duration, transiency, ratio of burst to total duration, frequency region with peak amplitude, and peak amplitude. Correlations between the values of the acoustic cues themselves are also listed (first 5 columns). Numbers in parentheses indicate the p-values for the associated correlation coefficients, and significant correlations are indicated by bold font.

	Duration	Transiency	Burst ratio	Frequency region	Amplitude	Percent correct
Duration	1	-.238 (.034)	-.068 (.550)	-.0165 (.144)	.225 (.045)	-.147 (.193)
Transiency	--	1	-.428 (<.001)	-.217 (.053)	.189 (.094)	.405 (<.001)
Burst ratio	--	--	1	.005 (.968)	-.271 (.015)	-.424 (<.001)
Frequency region	--	--	--	1	.066 (.559)	-.035 (.760)
Amplitude	--	--	--	--	1	.274 (.014)
Percent correct	--	--	--	--	--	1