Published in final edited form as:
J Acoust Soc Am. 2008 February ; 123(2): 1126–1139. doi:10.1121/1.2823453.

Perceptual learning of spectrally degraded speech and environmental sounds

Jeremy L. Loebach[a] and
Speech Research Laboratory, Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana 47405

David B. Pisoni[b]
Speech Research Laboratory, Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana 47405 and DeVault Otologic Research Laboratory, Department of Otolaryngology-Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, Indiana 46202

Abstract

Adaptation to the acoustic world following cochlear implantation does not typically include formal training or extensive audiological rehabilitation. Can cochlear implant (CI) users benefit from formal training, and if so, what type of training is best? This study used a pre-/posttest design to evaluate the efficacy of training and generalization of perceptual learning in normal hearing subjects listening to CI simulations (eight-channel sinewave vocoder). Five groups of subjects were trained on words (simple/complex), sentences (meaningful/anomalous), or environmental sounds, and then were tested using an open-set identification task. Subjects were trained on only one set of materials but were tested on all stimuli. All groups showed significant improvement due to training, which successfully generalized to some, but not all stimulus materials. For easier tasks, all types of training generalized equally well. For more difficult tasks, training specificity was observed. Training on speech did not generalize to the recognition of environmental sounds; however, explicit training on environmental sounds successfully generalized to speech. These data demonstrate that the perceptual learning of degraded speech is highly context dependent and the type of training and the specific stimulus materials that a subject experiences during perceptual learning has a substantial impact on generalization to new materials.

I. INTRODUCTION

Despite recent advances in cochlear implant technology, a large amount of variability in outcome and benefit is consistently reported among cochlear implant (CI) users that cannot be accounted for by differences in etiology, onset and duration of deafness, age at implantation and physiological factors (NIH, 1995). Given that there are no standardized rehabilitation programs consistently implemented after implantation, the experiences of CI users may differ from the start, placing them at fundamentally different baseline levels and contributing to the variability in the outcome measures. Could the standardization of training protocols establish a more stable baseline, and account for a portion of this variability? Moreover, what type of training is most effective, and yields the most robust levels of generalization to new materials? The present study was designed to assess the effectiveness of training when adapting to stimuli that have been processed by a cochlear implant

[a] Author to whom correspondence should be addressed. jlloebac@indiana.edu.
[b] pisoni@indiana.edu

simulation. Subjects were trained on speech (simple or complex single words; meaningful or anomalous sentences) or environmental sounds and compared on-open set recognition at posttest and during generalization to novel stimuli to quantify how explicit training affects the perceptual learning of stimuli processed by a CI simulation.

Cochlear implantation can provide sufficient acoustic input to a deaf individual to allow the establishment of some form of hearing (NIH, 1995). Whereas early implants provided the hope of recovering some auditory ability, most recipients of modern cochlear implants have the expectation that they will recover oral communication skills, including the ability to talk on the telephone (Shannon, 2005). In the worst case, patients are expected to regain some awareness of sound (Clark, 2002), including the detection and recognition of environmental sounds, however the degree to which CI users can actually recognize and identify environmental sounds is largely unknown (see, however, Reed and Delhorne, 2005).

Research using acoustic simulations of cochlear implants has met with great success (Shannon *et al.*, 1995). Vocoders simulate the limited number of spectral channels available in the electrode array by dividing the acoustic signal using a series of band-pass filters but preserve the temporal profile of electrical stimulation by modulating the noise bands with the amplitude envelope. Previous studies using the vocoder have demonstrated that successful speech recognition can occur in response to severely spectrally degraded stimuli; only a limited number of spectral channels are required so long as the temporal information in the envelope is preserved (Shannon *et al.*, 1995; Dorman *et al.*, 1997). Moreover, the performance of CI users on consonants and vowels was similar to that of normal hearing subjects listening to six-channel vocoded stimuli, demonstrating that the vocoder can successfully simulate the output of a CI in order to elicit equivalent levels of performance (Dorman and Loizou, 1998).

Although studies using vocoders have focused primarily on the identification of the linguistic content of the materials, the real world is composed of many other complex auditory events that are transmitted via the acoustic signal (Gaver, 1993). Compared to speech, considerably less is known about the perception of environmental sounds, both in the clear and processed by vocoders. Although there may be some commonalities between the perceptual systems required for the identification of speech and environmental sounds, the degree to which they operate independently is unknown. At a surface level, it appears that environmental sounds may be encoded in a similar manner to speech, in that the stimulus specific form may be preserved in addition to the more abstract symbolic lexical form (Lachs *et al.*, 2003) as demonstrated by cross-modal priming of environmental sounds (Chiu and Schacter, 1995; Chiu, 2000). More recently, the perception of both speech and environmental sounds has been shown to rely on a common auditory ability for familiar sound recognition (Kidd *et al.*, 2007). How efficiently a subject can locate stored information about an auditory stimulus, the problem solving strategies they engage in to constrain possible response options, and the ability to focus attention on the most important spectral and temporal information in the signal were all found to be common factors for the identification of environmental sounds and speech in noise (Kidd *et al.*, 2007). Moreover, the authors proposed a general auditory ability (*g*), which governs a listener's ability to process and perceive auditory information, and may be a necessary component for speech and language processing, as well as for identifying environmental sounds (Kidd *et al.*, 2007).

Additionally, a series of recent experiments by Gygi and colleagues demonstrated that the most important acoustic information for the recognition of environmental sounds overlaps with the information for speech (1200–2400 Hz, Gygi *et al.*, 2004). When processed with a vocoder, the results were much like those reported for speech: recognition accuracy increased with the number of channels (Gygi *et al.*, 2004). Moreover, the stimuli that showed the greatest improvement from training were those that had broader harmonic structure and spectral detail (Gygi *et al.*, 2004). Other work suggests that that the effects of processing environmental sounds using a noise vocoder may not be as straightforward as it is for speech (Shafiro, 2004). Closed-set recognition of environmental sounds improved as the number of channels increased, however, the improvement was stimulus dependent: environmental sounds that rely more on spectral information showed increases in accuracy with the addition of more spectral channels, whereas those that rely on temporal information showed decreases (Shafiro, 2004). Thus, it appears that some environmental sounds may show an altogether different pattern of spectrotemporal dependence as compared to speech signals.

Few studies have examined the perception of environmental sounds by CI users. Although CI users do improve in their ability to recognize environmental sounds following implantation, they do so at a slower rate than is typically observed for speech (Tye-Murray *et al.*, 1992). In a recent study, Reed and Delhorne (2005) demonstrated that CI users were significantly better at identifying environmental sounds (79% correct) than words (39% correct), and that subjects who were better at word identification were also better at environmental stimulus identification. Significant variability was observed across subjects, however, arising from differences in exposure to environmental sounds in their daily environment (Reed and Delhorne, 2005), and it is possible that additional exposure and explicit training could further increase performance.

One common theme throughout the studies using the vocoder is the issue of perceptual learning. Although subjects can accurately identify speech processed by a vocoder, a period of learning and adjustment is frequently required (Shannon *et al.*, 1995; Dorman *et al.*, 1997; Dorman and Loizou, 1998). Although some type of auditory training is necessary when adapting to acoustic simulations of cochlear implants, few studies have explicitly examined the effects of training and feedback on the adaptation to CI simulations. Moreover, the training conditions that maximize perceptual learning and promote robust generalization and transfer to new materials are not well understood. In a series of recent experiments, Davis and colleagues investigated the effects of training and feedback during adaptation to six-channel noise vocoded sentences (Davis *et al.*, 2005). When subjects were merely exposed to the stimuli, open set identification increased significantly across 30 sentences (a gain of 11%), which can be attributed solely to perceptual attunement to the synthesis conditions, since subjects received no feedback whatsoever. When subjects were provided with unprocessed auditory feedback, significant increases in recognition accuracy were also observed (Davis *et al.*, 2005). Although these results are encouraging, the typical CI user will not have access to the unprocessed version of a stimulus, making this type of feedback not clinically viable. Adding orthographic feedback proved to be just as effective as presentation of the original unprocessed acoustic version, producing robust gains over exposure alone (Davis *et al.*, 2005).

Another issue with perceptual learning lies in which stimulus materials are most effective during training. Davis and colleagues (2005) trained subjects to identify meaningful sentences, semantically anomalous sentences (sentences where the function words are intact, but the content words are unrelated), nonword sentences, or Jabberwocky sentences (anomalous sentences where the content words are replaced by nonwords). Although all groups showed significant improvement over the training interval, those who were trained on meaningful and anomalous sentences performed identically to one another, and significantly better than those trained on non-word and Jabberwocky sentences. These findings suggest that access to the syntactic structure of the materials may be required to elicit effective levels of perceptual learning, but when the content words are replaced with non-words the task of determining the syntactic structure becomes exceedingly difficult (Davis *et al.*, 2005).

As an initial investigation into the effectiveness of training and perceptual learning, the results reported by Davis and colleagues raise several important questions. Although they demonstrated that feedback has a significant impact on performance, the type of feedback they used would not necessarily be relevant for the typical CI user. An individual with electric hearing never has an opportunity to hear an unprocessed version of the stimulus. The finding that the subjects who received orthographic feedback paired with the vocoded version of the sentence performed just as well as those who received the unprocessed version, suggests that such feedback could be beneficial to individuals with cochlear implants. In addition, subjects who did not receive explicit feedback showed significantly lower levels of performance overall, but still achieved similar gains due to training. Due to methodological constraints, however, neither pre- to posttest gains in performance nor generalization to new materials could be assessed.

In another study, Burkholder (2005) demonstrated that the use of feedback consisting of the correct orthographic form of the sentence paired with the repetition of the vocoded stimulus produced significantly greater pre- to posttest gains than receiving the unprocessed version alone. Moreover, subjects who were trained on the anomalous sentences showed identical pre- to posttest gains to subjects trained on meaningful sentences, but showed significantly greater benefits during generalization to new materials including environmental sounds (Burkholder, 2005). These data suggest that access to the syntactic structure of the sentence without relying on sentence meaning may provide a greater benefit, presumably because the listener is forced to reallocate attention to the acoustic–phonetic structure of the signal and rely on bottom-up processes for recognition. Evidence showing that training successfully generalized to the identification of environmental sounds, underscores this point.

Several questions remain unanswered, however. Burkholder (2005) only assessed the generalization of training with speech to environmental sounds, but not the reverse. If subjects were truly relying on the acoustic structure of the stimuli, one would predict that training on environmental sounds should successfully generalize to speech. This possibility, however, has not been experimentally addressed. In addition, no baseline identification data were collected for the environmental sounds, so it is unclear if the subjects were performing significantly better at identifying the environmental sounds than with no training at all. Moreover, although training with meaningful sentences appears to generalize to novel sentences, it is unknown whether this training generalizes to single words. Anomalous sentences can be conceptualized as a series of unrelated words connected by a permissible syntactic structure. If this is the case, then training on single word identification should generalize to anomalous sentences, and training on anomalous sentences should generalize to single words. In addition, recent studies have shown that training on simple consonant vowel (CV) and consonant vowel consonant clusters (CVC) produces only modest gains in performance on sentence identification (Fu *et al.* 2006). It is unclear whether the converse is true; that is, would training on sentences, both high and low in context, generalize to single words and CVCs?

The purpose of the present study was to examine the effect of training on the recognition of speech and environmental sounds processed by a sinewave vocoder. Specifically, we assessed the perceptual learning of CVCs, words, meaningful sentences, anomalous sentences, and complex nonspeech environmental sounds using a pre-/posttest design, and then compared the generalization to different materials. As little is known on how exposure affects open-set recognition of environmental sounds, we collected baseline data from subjects who were exposed to but not explicitly trained on environmental sounds as a control. For sentences, mere exposure without feedback leads to gains of 11% over time (Davis *et al.*, 2005), suggesting that for both sentences and words, any gain from training that exceeds 10% is likely due to explicit training. Additionally, Burkholder (2005) demonstrated that subjects gain only 10% across three phases of training with semantically anomalous sentences when provided with unprocessed auditory feedback. Given that subjects were receiving feedback, it is expected that mere exposure to anomalous sentences would produce gains that would be far less than 10%.

II. METHODS

A. Subjects

One hundred fifty-five normal-hearing, healthy young adults from the Indiana University community participated in the study (107 female, 47 male, and 1 transgender). The age of the subjects ranged between 18 and 60, with a mean age of 22.38 years. All subjects reported that English was the first language that they learned in infancy. Most subjects (*n* = 141) were monolingual, but 14 reported being fluent bi- (*n* = 12) or trilingually (*n* = 2). Subjects were given credit in their introductory psychology course for their participation (*n* = 52), or were paid at the rate of $10 per hour (*n* = 103). Of these 155 subjects, five were excluded from the final data analysis, leaving 25 unique subjects in each training condition. One subject was excluded after reporting that they could not understand the stimuli as speech and one due to a program malfunction. After the experiment, one subject revealed that they were not a native English speaker, and so their data were excluded. Three subjects were excluded after the decision was made that they were not on task: one subject left many spaces intentionally blank and made frequent spelling errors that rendered the data impossible to score; one only transcribed the first few words of each sentence and it was thought that they had a possible memory deficit; and the final subject typed only gibberish (random keystrokes) rather than making a meaningful response to the stimuli.

B. Stimuli

Stimulus materials came from five different corpora that consisted of digital wav files of meaningful words, meaningful sentences, anomalous sentences, and environmental sounds.

1. Modified rhyme test—The modified rhyme test (MRT) corpus consisted of 300 words organized into 50 lists, where each list contains six rhymed variations on a common syllable (House *et al.*, 1965). Within each list, the word initial or word final consonant is systematically varied to produce six rhyming items (e.g., “bat,” “bad,” “back,” “bass,” “ban,” and “bath”). Stimuli consisted of 90 CVC words drawn from the MRT list, and their associated wav file recordings that were obtained from the PB/MRT Word Multi-Talker Speech Database in the Speech Research Laboratory at Indiana University Bloomington. Forty-two of the words were produced by a female talker, and the remaining forty-eight by a male talker.

2. Phonetically balanced words—The phonetically balanced (PB) corpus consisted of 20 lists of 50 monosyllabic words whose phonemic composition approximates the statistical occurrence in American English (e.g., “bought,” “cloud,” “wish,” and “scythe”) (Egan, 1948). Stimuli consisted of 90 unique words drawn from lists 1–3 of the PB corpus so that no overlaps occurred with those selected from the MRT corpus. Wav file recordings were obtained from the PB/MRT Word Multi-Talker Speech Database in the Speech Research Laboratory at Indiana University Bloomington. Half of the stimuli were produced by a male talker, and the other half by a female talker.

3. Harvard/IEEE sentences—The Harvard/IEEE sentence database consisted of 72 lists of 10 meaningful sentences (IEEE, 1969). These phonetically balanced (relative to American English) sentences contained five keywords embedded in a semantically rich meaningful sentence (e.g., "Her purse was full of useless trash," "The colt reared and threw the tall rider"). Stimuli consisted of 25 sentences drawn from lists 1–10 of the Harvard/IEEE sentence database and their associated wav file recordings obtained from the speech corpus originally created by Karl and Pisoni (1994) A female talker produced 14 sentences and a male talker produced the remaining 11. Selection of these two talkers was based on their production of speech that was highly intelligible (90% correct keyword accuracy across the 100 sentences) as demonstrated by previous research (Bradlow *et al.*, 1996).

4. Anomalous Harvard/IEEE sentences—Semantically anomalous sentences preserve the canonical syntactic structure of English, but have no meaning. Herman and Pisoni (2000) used the Harvard/IEEE sentence materials to create phonetically balanced semantically anomalous sentences. The keywords from the 100 sentences in lists 11–20 of the Harvard sentence corpus were coded according to semantic category (noun, verb, adjective, adverb) and replaced with words from equivalent semantic categories from lists 21–70 (Herman and Pisoni, 2000). This operation created sentences that have legal syntactic structure in American English, but were semantically anomalous (e.g., "Trout is straight and also writes brass," "The deep buckle walked the old crowd"), thus precluding subjects from using typical sentence context to identify the keywords. Stimuli consisted of 25 anomalous sentences drawn from the anomalous Harvard/IEEE sentence corpus of Herman and Pisoni (Herman and Pisoni, 2000) and their associated wav file recordings. A female talker produced 13 of the sentences, whereas a male talker produced the remaining 12 sentences.

5. Environmental sounds—The environmental sound database of Marcell and colleagues consists of acoustic signals recorded from a wide variety of acoustic environments developed for use in neuropsychological evaluation and confrontation naming studies (Marcell *et al.*, 2000). The corpus consists of 120 sounds from various acoustic events spanning a wide variety of categories: sounds produced by vehicles (e.g., automobile, airplane, motorcycle), animals (bird, dog, cow), insects (mosquito, crickets), nonspeech sounds produced by humans (snoring, crying, coughing), musical instruments (piano, trumpet, flute), tools (hammer, vehicles), liquids (water boiling, rain), among others. These sounds have been normed in a group of neurologically normal subjects on a variety of subjective (e.g., familiarity, complexity, pleasantness and duration) and perceptual measures (e.g., naming accuracy and naming response latency) (Marcell *et al.*, 2000). Stimuli consisted of ninety environmental sounds and their associated wav file recordings obtained from a digital database published by the authors on the Internet (http://ww.cofc.edu/~marcellm/confront.htm) (Marcell *et al.*, 2000). Stimulus selection from a variety of acoustic categories provided a wide representation of sound types and familiarity ratings.

C. Synthesis

Stimuli were processed using Tiger CIS (http://www.tigerspeech.com/) to simulate an eight-channel cochlear implant using the CIS processing strategy. Stimulus processing involved two phases, an analysis phase, which divides the signal into frequency bands and derives the amplitude envelope from each band, and a synthesis phase, which replaces the frequency content of each band with a sinusoid that is modulated with the appropriate amplitude envelope. Analysis used band-pass filters to divide the stimuli into eight spectral channels between 200 and 7000 Hz in steps with corner frequencies based on the Greenwood function (24 dB/octave slope). Envelope detection used a low pass filter with an upper cutoff at 400 Hz with a 24 dB/octave slope. Following the synthesis phase, the modulated sinusoids were combined and saved as 22 kHz 16 bit windows PCM wav files. Normalization of the wav files to a standard amplitude (65 dB rms) using a leveling program (LEVEL V2.0.3 Tice and Carrell, 1998) ensured that stimuli were equal in intensity across all materials, and that no peak clipping occurred.

D. Procedures

All methods and materials were approved by the Human Subjects Committee and Institutional Review Board at Indiana University. All subjects indicated their informed consent before beginning the experiment. A short subject information form asked for basic background information and inquired as to any prior hearing, speech, or language problems.

Data collection used a custom script written for PSYSCRIPT, and implemented on four Apple PowerMac G4 (512 Mb RAM) computers and four 15-in. color Sony LCD monitors (1024 × 768 pixels, 75 Hz refresh). Audio signals were presented over four sets of Beyer Dynamic DT-100 headphones, calibrated with a voltmeter to a 1000 Hz tone at 70 dB SPL using a voltage/intensity conversion table for the headphones. Sound intensity was fixed within PsyScript in order to guarantee consistent sound presentation across subjects.

Multiple booths in the testing room accommodated up to four subjects at the same time. Subjects were informed that the stimuli they would hear were processed by a computer and that while they may have difficulty understanding them at first, they would quickly adapt and that the purpose of the present study was to train them to understand the stimuli. On-screen instructions preceded each block to orient the subject to the materials and requirements of the upcoming task. Before the presentation of each audio signal, a fixation cross appeared at the center of the screen for 500 ms alerted the subject as to the upcoming trial and was followed by the presentation of the stimulus. Following stimulus offset, a dialog box appeared on the screen prompting subjects to type in what they heard. There were no time limits for responding; subjects performed at their own pace and were encouraged to rest between each trial as needed. The experimental session lasted on average 45 min. All subjects received written and verbal debriefing after the experiment.

1. Training—Each training condition consisted of seven blocks. Stimuli were prerandomized and organized into separate lists for presentation in each training condition. Although the stimuli used in each block varied as a function of training materials, the same basic block design was consistent throughout all conditions (Table I).

In order to establish a baseline level of performance before training, a pretest was conducted in which subjects transcribed stimuli from the appropriate training category, but did not receive any feedback. During training, subjects received feedback in the form of the repetition of the processed auditory stimulus paired with the written form of the stimulus on the computer screen irrespective of whether their previous response was correct (the transcription of the word or sentence, or the descriptive label of the environmental sounds). During the posttest, subjects heard a selection of old materials from the pretest and training, as well as new materials from the same category. The posttest materials were selected to assess the effects of explicit training (using training materials), familiarity without explicit training (pretest materials) and novelty (previously unheard materials). The remaining three blocks assessed the generalization of training to other types of materials.

a. MRT word training: During the pretest, listeners were presented with 20 MRT words. Training consisted of 50 novel MRT words. An intervening generalization block occurred in block 3 to prevent habituation to the stimuli, and consisted of 25 anomalous sentences. The posttest in block 4 presented a total of 60 MRT words, 20 of which were drawn from the pretest materials, 20 from training, and 20 were novel stimuli with which subjects had no previous experience in the experiment. The remaining three blocks tested generalization to 25 meaningful sentences (block 5), 50 PB words (block 6), and 60 environmental sounds (block 7).

b. PB word training: PB training utilized an identical design to the MRT training, except the pretest, training and posttest materials consisted of PB words and block 6 tested generalization to 50 novel MRT words.

c. Anomalous sentence training: In order to allow for the relative effect of words transcribed across sentences, fewer sentences were selected. The pretest block consisted of four anomalous sentences (20 key words); the training block consisted of ten novel anomalous sentences (50 key words). Block 3 was an intervening block, consisting of 50 MRT words. The posttest in block 4 utilized 12 anomalous sentences, four selected from the pretest, four from the posttest, and four novel sentences (60 keywords). The remaining three blocks tested the effects of generalization to new materials. Block 5 consisted of 25 meaningful sentences, block 6 of 50 PB words and block 7 of 60 environmental sounds.

d. Harvard/IEEE sentence training: Harvard/IEEE sentence training utilized an identical design to the anomalous sentence training, except that the pretest, training and post-test materials consisted of meaningful sentences, and block 6 tested generalization to 25 anomalous sentences.

e. Environmental sound training: Like the MRT and PB training, training on environmental sounds began with a pretest consisting of 20 environmental sounds and training consisting of 50 novel environmental sounds. Subjects were asked to respond by describing what event or object produced the sound and were given a typical example that did not occur in the materials. Subjects were told that if they heard something that sounded like music they should try to indicate what musical instrument produced the sound, rather than simply identifying it as music. An intervening block occurred in block 3 in order to prevent habituation to the stimuli and consisted of 25 anomalous sentences. The posttest in block 4 presented a total of 60 environmental sounds, 20 of which were drawn from the pretest materials, 20 from training, and 20 were novel stimuli with which subjects had no previous experience in the experiment. The remaining three blocks consisted of generalization to 50 MRT words (block 5), 25 Harvard/IEEE sentences (block 6), and 50 novel PB words (block 7).

A separate group of 25 subjects served as controls in the environmental stimulus training condition. These subjects were exposed to the same processed training stimuli as the experimental group, but did not receive any feedback. Pre- to posttest comparisons were carried out to determine whether subjects showed any gains from being exposed to the stimuli and processing conditions during training, and were compared to subjects in the experimental group to assess the effectiveness of explicit training on environmental sounds. Pre- and posttest scores were also compared to environmental stimulus performance of the other training groups (HS, AS, MRT, and PB) to further assess whether training on speech provided any benefit on the recognition of environmental sounds.

2. Analysis and scoring—A supervised spellchecker in Microsoft EXCEL was used to correct the more obvious spelling errors and standardized responses across subjects by recoding homophones into a standard spelling. An automated macro searched for target/response matches using a preordained target list, the result of which was hand checked by a research assistant. Responses that were morphologically related to the target (pluralization of nouns or conjugation of verbs) were scored as incorrect. PB and MRT words were scored based on whether the entire word was correct, whereas the anomalous and meaningful sentences were scored for keywords correct (five keywords per sentence).

Environmental sounds were checked using a similar procedure. Scoring rules were modified slightly from those originally used by Marcell and colleagues (2000) given the nature of the signal degradation. Responses were scored as correct if the subject identified the target agent (e.g., cow), the sound the agent made if it did not have multiple possible agents (e.g., moo), or the linking of the two (e.g., cow mooing) (Gaver, 1993). Failure to specify agent, or incorrectly specifying agent was scored as an incorrect response. Correct identification of musical instruments required accurate identification of the instrument. The generic response of "music" was scored as incorrect, given that the instructions explicitly told subjects that this was not a valid response option.

For each training condition, responses were averaged across subjects for each block. Within-subjects analyses compared performance across blocks of a given training condition. Paired samples t-tests were used to assess the effects of training by comparing pre- and posttest performance. Posttest scores were balanced by only including the responses to the materials on which subjects were not explicitly trained, to avoid biasing the findings: none of the posttest values reported in the text, tables, or figures contains responses to the stimuli used during training. The differences in performance on the source of the posttest materials (items from pretest, training, and novel lists) were assessed with a one-way analysis of variance (ANOVA) and post hoc Tukey tests which were corrected for multiple comparisons using the Bonferroni correction. Other paired t-tests were conducted to assess the effects of context (anomalous sentences versus Harvard/IEEE sentences) and complexity (PB words versus MRT words). A correlational analysis examined the relationship between performance across blocks to assess whether performance on one set of materials was correlated with performance on another. Between subjects comparisons assessed the effect of training on materials across training conditions using one-way ANOVA and post-hoc Tukey tests.

III. RESULTS

A. Pre-/posttest comparisons

All types of training produced significant pre- to posttest gains in performance (Table II). Moreover, all speech materials showed a benefit from training that was greater than 10%, indicating that explicit training produced changes that exceeded the effects of mere exposure reported by Davis and colleagues (2005).

For the MRT stimuli, subjects increased significantly from 5.8% correct at pretest to 37.5% after training [$t(1,24) = 13.576$, $p < 0.001$]. A one-way ANOVA comparing the source of the posttest materials demonstrated that subjects performed significantly better on materials on which they were explicitly trained (58% correct) than on materials with which they were familiarized but not explicitly trained (40.4% correct) and novel materials (34.6% correct) [$F(2,72) = 18.967$, $p < 0.001$]. Post-hoc Tukey tests revealed that subjects performed significantly better on posttest stimuli drawn from the training list (both $p < 0.001$) demonstrating a significant effect of feedback, and indicating good retention of training by subjects. However, subject performance did not differ on post-test materials drawn from the pretest and novel lists ($p = 0.313$).

Subjects trained on the PB words started out better than those subjects trained on the MRT words, also showing significant increases from pre- (23.4% correct) to posttest (46.2% correct) [$t(1,24) = 7.134$, $p < 0.001$]. Examination of the posttest materials revealed that subjects performed best on stimuli on which they were explicitly trained (55.4% correct), followed by novel PB words (48.20% correct) and words on which they were previously exposed, but not explicitly trained (44.20% correct) [$F(2,72) = 5.484$, $p = 0.006$]. Post-hoc Tukey tests revealed that subjects performed significantly better on the post-test materials drawn from the training list ($p = 0.005$) than on materials from the pre-test list, but no difference was observed for materials drawn from the novel list ($p = 0.097$). More importantly, subject performance did not differ between materials drawn from the pretest and novel list ($p = 0.477$).

For the anomalous sentences, performance was good at pretest (33.6% correct) but increased significantly following training (61.7% correct, $t(1,24) = 11.713$, $p < 0.001$). Examination of the posttest materials revealed a significant main effect of source [$F(2,72) = 14.115$, $p < 0.001$], and post-hoc Tukey tests confirmed that subjects performed significantly better on the posttest materials drawn from the training list (78.2% correct) than on materials from either the pretest (61.6% correct, $p < 0.001$) or novel list (61.8% correct, $p < 0.001$). No differences in performance were observed on the materials from the pre-test and novel lists ($p = 0.998$).

Performance on the Harvard/IEEE sentences also increased significantly from pre- (40% correct) to posttest (63.9% correct, $t(1,24) = 7.041$, $p < 0.001$). A one way ANOVA revealed a significant main effect of source [$F(2,72) = 114.043$, $p < 0.001$] indicating that subjects performed significantly better on materials from the training list (97% correct) than on those from the pretest (71.6% correct) and novel (56.2% correct) lists (all $p < 0.001$). Subjects also performed significantly better on the post-test materials drawn from the pre-test list as compared to the novel list ($p < 0.001$). This is likely due to the high contextual salience of the sentences, because this pattern was not observed for the anomalous sentence training group.

Performance on the environmental sounds also showed a significant benefit from explicit training (Fig. 1). Subjects in the experimental group showed significant improvement between pre- (38.2% correct) and posttest [46.4% correct, $t(1,24) = 2.804$, $p = 0.01$]. An analysis of the posttest materials revealed a significant main effect of source [$F(2,72) = 8.717$, $p < 0.001$]. Subjects performed best on stimuli from the novel list (53.2% correct), followed by materials from the training list (50% correct) and pretest (39.6% correct). Subjects performed significantly better on posttest materials from both novel and training lists than on materials from the pretest list ($p = 0.009$ and $p < 0.001$, respectively) but did not differ from one another ($p = 0.617$).

Subjects in the environmental sound control group did not show improvement between pre- (39% correct) and posttest [37% correct, $t(1,24) = 1.056$, $p = 0.301$]. A univariate ANOVA comparing performance on the pre- and posttest blocks across the control and experimental groups revealed a significant main effect of training [$F(3,96) = 5.154$, $p = 0.002$]. Subjects in both groups performed equally well during the pretest phases ($p = 0.991$), but at posttest subjects in the experimental group (46.4% correct) performed significantly better than subjects in the control group (37% correct, $p = 0.003$). These data indicate that subjects who received explicit training on environmental sounds show gains in performance above and beyond the gains obtained from merely being exposed to the stimuli.

B. Correlations

Correlations of the performance across blocks revealed several interesting results (Table III). In general, the specific type of materials that subjects received in training was most strongly correlated with materials of a similar class (words with words, sentences with sentences).

For the MRT words, performance at posttest was most strongly correlated with performance on the PB words ($r = 0.766$, $p < 0.001$) followed by Harvard/IEEE sentences, and anomalous sentences, but not environmental sounds. Moreover, similar relationships were observed for the PB words and meaningful and anomalous sentences. It is interesting to note that performance on isolated words was most strongly correlated with performance on other words, followed by meaningful and anomalous sentences, and that sentences were most strongly correlated with other sentences.

Performance on the PB word posttest was most strongly correlated with performance on the MRT words followed by Harvard/IEEE sentences, but was not correlated with anomalous sentences or environmental sounds. MRT performance was also correlated with performance on anomalous sentences and environmental sounds. As observed in the MRT training group, performance on words (PB or MRT) was most strongly correlated with performance on other words and performance on sentences was most strongly correlated with performance on other sentences.

Performance on the anomalous sentence posttest was only correlated with performance on Harvard/IEEE sentences ($r = 0.63$, $p = 0.001$). The only other significant correlation observed was between PB words and environmental sounds ($r = 0.446$, $p = 0.025$). All other correlations were not significant.

Performance on the Harvard/IEEE sentence posttest was most strongly correlated with performance on anomalous sentences, followed by PB and MRT words. Anomalous sentences were most strongly correlated with performance on PB and MRT words.

Performance on the environmental stimulus post-test was not significantly correlated with any of the other materials, but as observed earlier, Harvard/IEEE sentences were significantly correlated with anomalous sentences.

C. Across Group Comparisons

To assess the effect of training on the source materials, the recognition accuracy scores for a given set of materials were compared across training conditions (Fig. 2). Comparison with the pretest data assessed whether the type of training a subject received had an effect on performance. Comparison with the posttest scores assessed whether the type of training significantly affected performance, and whether training on a specific set of materials produces better and more robust generalization than another.

Examination of the performance on the MRT words across training materials [Fig. 2(a)] with a one-way ANOVA revealed a significant main effect of training materials [$F(5,144) = 37.495$, $p < 0.001$]. Post-hoc Tukey tests revealed that subjects performed significantly better than the pretest regardless of the type of material that they were trained upon (all $p < 0.001$). This is not surprising, given the poor baseline performance (5.8% correct). Although any type of training produced a benefit, subjects trained on single words regardless of their origin (MRT and PB) performed identically ($p = 0.477$) and significantly better than subjects trained on any other materials (all $p < 0.01$). Training on anomalous sentences, Harvard/IEEE sentences and environmental sounds also produced significant gains over baseline, performing similarly on the MRT generalization test (all $p > 0.319$). When the scores were grouped by material type, however, subjects who received training on words (MRT and PB) performed significantly better than subjects trained on sentences ($p < 0.001$) or environmental sounds ($p = 0.027$).

Training also produced a significant main effect on performance on the PB materials [$F(5,144) = 24.86$, $p < 0.001$; Fig. 2(b)]. Overall, it did not matter what type of training subjects received, as performance was significantly higher than pretest for all training conditions (MRT training 43.44% correct $p < 0.001$, AS training 44.08% correct $p < 0.001$, HS training 44.08% correct $p < 0.001$, ENV training 43.68% correct $p < 0.001$). The main effect for training condition is carried entirely by the gains in performance relative to the pretest, because no significant differences were observed between performance across the five training conditions (all $p > 0.867$). This pattern indicates that when identifying words that are highly discriminable, training with any type of material will provide an equivalent benefit.

A one-way ANOVA on the anomalous sentences [Fig. 2(c)] revealed a significant main effect of training [$F(5,144) = 22.986$, $p < 0.001$]. Comparison with the pretest revealed that all types of training produced significant increases in performance relative to the baseline (33.6% correct, all $p < 0.001$). Subjects who were trained the anomalous sentences (61.7% correct) performed as well as those who were trained on the meaningful Harvard/IEEE sentences (58.62% correct, $p = 0.902$) and significantly better than those trained on PB, MRT, and environmental sounds (all $p < 0.004$) sentences. Training on MRT (47.74% correct), PB (46.53% correct) and environmental sounds (47.74% correct) provided equivalent benefit when recognizing the anomalous sentences, however (all $p < 0.0998$).

A significant main effect of training condition was also observed for the Harvard/IEEE sentences [$F(5,144) = 22.444$, $p < 0.001$; Fig. 2(d)]. The comparison of each of the training conditions to the Harvard/IEEE sentence pretest revealed that all subjects performed significantly better than the baseline (40% correct) regardless of the type of training they received (all $p < 0.001$). As was the case for the PB materials, the training effect was carried entirely by the gains in performance relative to the pretest, as there were no significant differences between performance across the five training conditions (MRT 67.01% correct, PB 66.21% correct, HS 63.90% correct, AS 67.04% correct, ENV 68.51% correct, all $p > 0.719$).

The effect of training on the recognition of the environmental sounds showed a different pattern of results [Fig. 2(e)]. A one-way ANOVA revealed a significant main effect of training group on performance [$F(7,192) = 4.452$, $p < 0.001$], however unlike the training effects observed for the other stimulus materials, subjects only showed gains relative to baseline (38.2% correct) when they were explicitly trained on the environmental sounds (46.40% correct, $p = 0.013$). Subjects trained on all other materials failed to show any differences as compared to baseline (MRT 37.6% correct $p = 1.00$; PB 35.2% correct $p = 0.822$; HS 34.93% correct $p = 0.999$; AS 34.93% correct $p = 0.764$). Pre- (39%) and posttest (37%) performance of subjects in the control group did not differ from subjects in the AS, HS, PB, or MRT training groups (all $p > 0.7$) indicating that subjects who were exposed to but not explicitly trained on the environmental sounds perform as well as subjects who are trained on speech. In addition, explicit training on environmental sounds produced significantly better posttest performance compared to all other groups (all $p < 0.03$), who did not differ from one another (all $p > 0.557$). That these scores did not differ from the baseline (for either the experimental or control groups), however, suggests that training on the speech materials was equally ineffective when transferring to environmental sounds. In effect, when asked to identify environmental sounds, training with speech materials is as effective as not receiving any training at all.

One potential concern with the current study is the use of a blocked design, which could lead to order effects at testing. Although all groups performed the posttest in block 4, the generalization materials were presented in different blocks to each group due to the block design. To assess the potential order effect a univariate ANOVA was conducted on each type of material using presentation order as the between subjects variable and training as the covariate. Although an order effect was observed for the MRT words [$F(2,122) = 10.185$, $p < 0.001$], post-hoc Tukey tests revealed that the only differences between blocks was between posttest (block 4) and when the MRT stimuli were presented in block 3 (Harvard/IEEE and anomalous sentence training, $p = 0.002$). This result may be more apparent than real, because performance on MRT words in block 3 did not differ from performance in blocks 5 or 6 ($p = 0.122$, $p = 1.00$) respectively, suggesting that additional practice effects may not have significantly influenced performance on the MRT words during generalization.

For the rest of the generalization data, order effects were either not observed, or could be accounted for entirely by training. For PB words, tests of generalization always occurred in block 6, and no significant effect of order was observed [$F(2,122) = 0.975$, $p = 0.325$]. Similar results were observed for Harvard/IEEE sentences [$F(2,122) = 1.003$, $p = 0.37$] in which generalization was tested in blocks 5 or 6. For anomalous sentences, generalization always occurred in block 5, and a significant main effect of order was observed [$F(2,122) = 22.77$, $p < 0.001$]. A post-hoc Tukey test revealed that this effect was entirely confounded with training, because subjects performed significantly better in the posttest (block 4) than during generalization (block 5, $p < 0.001$). Generalization to environmental sounds also showed a significant main effect of order [$F(2,122) = 24.650$, $p < 0.001$], but the effect was confounded with training, as post-test performance in block 4 was significantly higher than generalization in block 7 ($p < 0.001$). Thus for all data order effects were not apparent, or were completely accounted for by training. For MRT words a potential order effect was indicated, but since performance did not differ between early and late blocks, this effect may be more apparent than real.

The specific gains from training are displayed in Fig. 3. Gain scores were computed by subtracting performance at pretest from performance during generalization and posttest for all materials. Positive gain scores indicate improvement after training. Overall, the largest gains from training were observed for the MRT words [Fig. 3(a)], for which training on either MRT words or PB words produced significantly higher gains from training than either type of sentences or environmental sounds. A similar result was observed for anomalous sentences [Fig. 3(c)], where training on either anomalous or meaningful sentences produced significantly higher gains than for words or environmental sounds. Gains for PB words [Fig. 3(b)] and meaningful Harvard/IEEE sentences [Fig 3(d)] were uniform across all training conditions. Finally, training on environmental sounds [Fig. 3(e)] was the only form of training that produced a significant benefit in the identification of environmental sounds.

IV. DISCUSSION

Taken together, our results showed that the specific type of stimulus materials used during training had a significant effect on perceptual learning. Subjects showed significant pre- to posttest improvement in all training conditions, demonstrating that they were able to make effective use of feedback to improve their performance. Generalization effects, however, were not uniform across materials. For some training materials, subjects showed encoding specificity, performing best on the materials on which they were explicitly trained (e.g., Tulving and Thomson, 1973). Subjects who were trained on isolated words (PB or MRT) performed significantly better when identifying MRT stimuli than the other groups. When tested on PB words, all subjects regardless of training, performed equally well. Subjects who were trained on sentences (anomalous or meaningful) performed significantly better when identifying anomalous sentences than the other groups. When tested on Harvard/IEEE sentences, all subjects performed equally well regardless of the materials on which they were trained. The differences in training specificity suggest that differences exist in the difficulty of the materials, with training generalizing more readily to easier tasks (as evidenced by the nonspecific effect of training on generalization to PB words and Harvard/ IEEE sentences), but not to more difficult tasks (as evidenced by training specificity that was observed for MRT words, anomalous sentences and environmental sounds).

Although MRT and PB words have similar word frequencies in American English, they come from lexical neighborhoods with different densities. Neighborhood density determines how many confusable "neighbors" a given word has based on phonological similarity (how many different words can be formed by changing a single phone: beach, peach, teach, breach, bleach, etc.) (Luce and Pisoni, 1998). Words from sparse neighborhoods have fewer confusable lexical alternatives and are recognized faster and with higher accuracy than words from more dense neighborhoods (Luce and Pisoni, 1998). The neighborhood density and lexical frequency for the PB, MRT, and meaningful and anomalous sentence keywords were obtained from the Neighborhood Activation Model database (http://neighborhoodsearch.wustl.edu/neighborhood/Home.asp). PB words had a lower average neighborhood density (12.25) compared to MRT words (21.80) suggesting that PB words should be more discriminable and identifiable than MRT words. As anomalous sentences contain keywords that were initially drawn from the Harvard/IEEE sentences they are equal in lexical frequency and neighborhood density (10.83 and 10.72, respectively). The main difference between these sentences is semantic predictability. The differences in the predictability and confusability of the speech stimuli likely produced differences in task difficulty. When task demands were high (due to lower predictability or higher confusability), subjects performed better when they were trained on stimuli of the same general class (e.g., training on words generalized significantly better to other words, sentences generalized significantly better to other sentences), demonstrating transfer appropriate processing (e.g., Morris, *et al.*, 1977). The opposite effect was observed for the more predictable materials: subject performance did not differ across training groups on the PB words and Harvard/IEEE sentences. This suggests that when the task demands are less difficult, such as when identifying words from lower density neighborhoods or highly predictable sentences, all forms of training are equivalent.

It is interesting to note here that training on PB words had a larger effect on the recognition of MRT words (*M* = 47%) than on the PB words themselves (*M* = 43%) [Fig. 3(a) compared to Fig. 3(b)]. This is likely due to the poorer performance during the pretest for MRT words (*M* = 6%) relative to PB words (*M* = 23%): subjects had more of an opportunity to improve their performance on the MRT words even though they were being trained on PB words. The differences in the performance at pre-test likely stem from the differences in task difficulty: MRT words are composed of syllables that differ only on word initial or word final consonants and are from higher density lexical neighborhoods, whereas PB words are phonetically balanced for phoneme occurrence in American English and come from lower density lexical neighborhoods. For the MRT words, all groups did improve significantly relative to the low baseline, but training on MRT and PB words had an additional effect, improving performance significantly more than either of the sentence training conditions or the environmental sound training condition. Therefore, the low baseline for the MRT words at pretest cannot account for the differences in performance, and the higher performance of subjects trained on MRT or PB words must be due to training.

Although training on words and environmental sounds generalized to anomalous sentences, it is unknown how much exposure to anomalous sentences without explicit training or feedback would improve performance. Based on the previous research of Davis and colleagues (2005), exposure to vocoded meaningful sentences without feedback only produced gains in performance of 11%. It is likely that much of this effect is driven by the semantic predictability of the materials, and that improvement from exposure to anomalous sentences in which sentence context cannot be used to predict upcoming words, would be much lower. Additionally, Burkholder (2005) demonstrated that giving subjects unprocessed auditory feedback during training on vocoded anomalous sentences improved performance across three phases of training by 10%. Given that these subjects were provided with feedback, we would expect the gains from exposure to anomalous sentences to be far less than 10%. In the present study, subjects trained on MRT and PB words and environmental sounds showed improvement on the recognition of anomalous sentences that exceeded 10% over the anomalous sentence pre-test (gains of 14%, 13%, and 16%, respectively), suggesting that such training improved performance on anomalous sentences that was greater than would be expected from exposure alone. Future work, however, will have to determine how much of an effect exposure has on the recognition of vocoded anomalous sentences.

Overall, the specific type of materials upon which subjects were trained was most strongly correlated with materials of a similar class. For both MRT and PB training, posttest performance was most strongly correlated with performance on the PB/MRT generalization blocks, respectively. For anomalous and Harvard/IEEE training groups, posttest performance was most strongly correlated with performance on the Harvard/IEEE and anomalous sentence generalization blocks, respectively. These findings provide further support for encoding specificity and transfer appropriate processing. Except for three instances, performance on speech stimuli was not correlated with performance on environmental sounds. Performance on MRT and PB materials was significantly correlated with performance on the environmental sounds for the PB training, anomalous and Harvard/IEEE training, suggesting some bidirectionality to training on speech. Interestingly, environmental stimulus training was not correlated with performance on any tests of speech generalization.

One intriguing finding from this study was the clear asymmetry in training that was observed for the environmental sounds (Fig. 3). Subjects explicitly trained on environmental sounds performed significantly better than baseline on all speech materials, suggesting that training on complex nonspeech stimuli produces robust generalization to speech. The inverse, however, was never observed: training on speech consistently failed to produce performance that differed from the environmental baseline. Compared to the control group, explicit training had a significant effect on posttest recognition of environmental sounds, indicating that explicit training generalized to new environmental stimuli that subjects had not heard previously in the experiment. In addition, subjects who were trained on speech did not differ from subjects in the control group who were exposed to the environmental sounds but did not receive explicit feedback or training. Thus, it appears that explicit training on complex non-speech materials leads to improved performance on speech materials, but training on speech materials does not produce gains in the perception of complex non-speech abilities. This suggests that explicit environmental sound training may have increased subjects' attentional sensitivity to the spectrotemporal characteristics of the stimuli, which may have enhanced their utilization of similar information for the identification of speech. However, training on speech may predispose the listener to make lexical judgments about the signal, placing them in a processing mode that focuses on stimulus meaning, a higher order cognitive variable, and reducing their attention to the lower order acoustic features. Further research will be necessary to determine the extent to which such bidirectionality exists.

Recent neuroimaging studies investigating the encoding of environmental sounds have suggested that similar cortical regions may be involved during the processing of environmental sounds and speech sounds (Lewis *et al.*, 2004). These cortical regions include canonical auditory areas required for the recognition of sound (primary auditory cortex), the identification of auditory speech stimuli (superior temporal gyrus, posterior superior temporal sulcus), semantic processing and accessing of lexical information during sound, picture and action naming (posterior medial temporal gyrus, pMTG) (Lewis *et al.*, 2004). These cortical areas (the pMTG and pSTS in particular) showed bilateral activation in response to environmental sounds, but tend to be left lateralized during speech perception tasks (Lewis *et al.*, 2004). This difference may partially explain the asymmetry that we observed for training with environmental sounds and speech. Perhaps training with environmental sounds activated cortical regions implicated in the processing of speech stimuli, leading to efficient generalization to speech. Due to different task demands, training with speech may have utilized additional lateralized cortical regions that would not necessarily facilitate generalization to environmental sounds.

Additionally, other recent neuroimaging studies have demonstrated that the functional connectivity between cortical regions may be differentially altered due to task demands when identifying speech (Obleser *et al.*, 2007). This may facilitate generalization in one case (environmental sounds to speech), but not the other (speech to environmental sounds). Taken together with the recent findings of Kidd and colleagues (2007), a general auditory ability for the recognition of familiar sounds may contribute to the identification of both speech and environmental sounds, providing further support for using general auditory training to enhance the perceptual abilities of CI users.

The present findings are similar to those of Gygi and colleagues, who found that the most important information for recognition of environmental sounds occupies an identical frequency range as that for speech (Gygi *et al.*, 2004). If the important information for environmental sounds overlaps substantially with speech, then training subjects to better utilize the spectro-temporal information in this frequency region more efficiently should improve generalization to speech, as we report here. Training on speech alone may not be sufficient to increase generalization to environmental sounds, because subjects may focus their attention on the higher order lexical and semantic attributes of the signal rather than the spectrotemporal information itself. The finding that training on speech does not generalize to environmental sounds conflicts with the earlier findings of Burkholder (2005), who reported that training on speech did generalize to environmental sounds. However, Burkholder did not use a pre-/posttest design, so the baseline performance levels for environmental sounds were not known. Although training on speech materials in the present study led to performance levels for environmental sounds that were greater than zero, they did not exceed the pretraining baseline. This result suggests that subjects in the Burkholder study (2005) may not have performed any differently after training than subjects who were naïve to the stimulus processing conditions.

Surprenant and Watson (2001), and more recently Kidd and colleagues (2007) reported a significant correlation between subject’s ability to discriminate nonspeech stimuli based on spectrotemporal cues and their identification of speech in noise. The authors suggested that common higher order acoustic processes may contribute to both speech and nonspeech processing capabilities. These differences could account for the substantial differences in performance of subjects who receive hearing aids, and cochlear implants alike: auditory sensitivity at a peripheral level may not be the sole cause of variability in outcome and benefit; rather the inability to utilize and manipulate auditory information at higher levels may supersede the benefits of an auditory prosthesis (Surprenant and Watson, 2001). This relationship may not be completely bidirectional, however, given our findings that training on environmental sounds generalizes to speech, but training on speech does not generalize to environmental sounds.

One important methodological difference between the results of the present study and earlier investigations using environmental sounds is the use of open-set testing procedures in all conditions. Gygi and colleagues reported identification scores of up to 66% correct using 6-channel noise vocoded stimuli (Gygi *et al.*, 2004), and Shafiro found that performance reaches asymptote with 16 channels (66%), but large stimulus specific effects were observed (Shafiro, 2004). Moreover, Reed and Delhorne (2005) found that CI users show higher levels of performance (79% correct). One commonality between all three earlier studies is that testing used closed-set forced-choice procedures. Under open-set testing, average performance after training (46% correct) was substantially lower than the performance observed in the previous studies. Given that the closed-set procedures necessarily limit subjects to a certain set of responses, open-set testing allows subjects to record their actual impressions of the stimuli in a way that would be more appropriate to real world listening environments (see Clopper *et al.*, 2006).

If subjects are indeed learning to utilize the residual spectrotemporal information more efficiently during training on environmental sounds, we should expect to see benefits for CI users as well. One problem, however, is that not all environmental sounds are perceptually equivalent. As Shafiro (2004) and Burkholder (2005) noted, some environmental sounds may be inherently more identifiable than others. The trading relations between the number of spectral bands needed for environmental sound recognition found by Shafiro (2004) suggests that some stimuli may not be readily identifiable when processed by a vocoder. Given that the amount of acoustic information differs across acoustic environments and task demands, the spectral resolution of the current generation of cochlear implants may be insufficient to provide significant benefit under all listening situations (Shannon *et al.*, 2004; Shannon, 2005). Training CI users to better make perceptual distinctions based on the acoustic information that they do have may provide additional benefit when listening to nonspeech sounds or identifying speech in noise.

Further, an important theoretical question is also raised here. Although several previous studies have failed to show substantive differences for the perception of speech as processed by a noise and sinewave vocoder (Dorman *et al.*, 1997), other studies have found that for non-speech tasks, performance is actually better for sinewave vocoded speech (Gonzales and Oliver, 2005). Gender and talker identification were significantly better for stimuli processed using a sinewave vocoder than when processed using a noise vocoder (Gonzales and Oliver, 2005). The authors suggest that the sinewave carriers may have introduced less distortion, thus preserving more accurate and robust detail in the amplitude envelopes that could be useful to the listener. A comparison of the two methods revealed more residual periodic information in the sinewave vocoder processed signal as compared to the noise vocoder processed signal, forming the basis for their claim (Gonzales and Oliver, 2005). It may be the case that a sinewave vocoder may produce better, more robust results for studies using music and environmental sounds than would a noise vocoder: for stimuli that carry more salient spectral information, less distortion and better preserved periodicities in the envelope may translate to heightened recognition. Whether performance on these types of stimuli differ from performance of cochlear implant users remains an open question.

Our findings also replicate and extend the recent studies conducted by Davis and colleagues (2005) and Burkholder (2005). Training using orthographic feedback paired with a repetition of the processed version of the sentence produced keyword identification scores (71% correct) that were nearly identical to those reported by Davis in the last block of training (75% correct). We also found that training on anomalous sentences produced excellent generalization to meaningful sentences, as was reported previously by both Davis *et al.* (2005) and Burkholder (2005). Moreover, the gains from training exceeded 11%, which was the benefit observed by Davis and colleagues (2005) for exposure to the stimuli without feedback. Thus, access to syntactic structure without relying on sentence context enhances sentence recognition. Our extension to include single PB words and CVCs also provides additional support for this conclusion: training on all materials produced excellent generalization to the meaningful Harvard/IEEE sentences. The results observed for training on environmental sounds, however, suggest that learning to recognize the acoustic form of a stimulus enhanced selective attention to spectro-temporal information, and bottom-up perceptual encoding processes.

We also replicated the findings of Fu and colleagues, who showed that giving CI users explicit training on CV and CVCs does indeed produce gains in sentence intelligibility (Fu *et al.*, 2006). The similar patterns of performance observed with normal hearing subjects listening to acoustic simulations of a cochlear implant provides further support for the utility of the vocoder as an effective model of electric hearing. By studying the perceptual learning of CI simulated speech in normal hearing listeners we can simultaneously learn about the neural and cognitive mechanisms that underlie speech and language processing in general, and expand our knowledge about effective rehabilitation and training programs to assist newly implanted individuals. By formalizing training paradigms that utilize a wide variety of stimulus materials, we may be able to provide cochlear implant users with tools that will bootstrap onto a variety of tasks and difficult listening conditions above and beyond those on which they were trained (i.e., increase “carry-over” effects). Given the substantial variability in performance among cochlear implant users that cannot be attributed to individual differences in etiology and duration of deafness, the question remains as to how differences in postimplantation experience contribute to outcome and benefit. Providing explicit instruction as to the important information in the signal may help to account for a portion of this variability, thereby allowing us to disentangle the role of experience and provide a more objective assessment of cochlear implant user success.

In summary, we demonstrated that the type of stimulus materials used during training affects generalization to new materials. Although all forms of training provided some benefit, generalization of training was not uniform, and was highly context and task specific. When the task was easy, such as was the case when identifying contextually rich meaningful sentences or highly discriminable isolated words, all five training conditions provided equivalent benefits. When the task was more difficult, such as was the case when identifying highly confusable CVCs or sentences lacking semantic context, subjects who were trained on materials of a similar nature to those on which they were being tested performed significantly better. However, the addition of environmental sounds revealed a unique asymmetry: training on environmental sounds generalized to the recognition of speech, but training on speech did not generalize to environmental sounds. This pattern of performance suggests that a wide variety of stimulus materials should be used during training to maximize perceptual learning and promote robust generalization to novel acoustic sounds.

Acknowledgments

The authors would like to thank Dr. Qian-Ji Fu for developing and maintaining TigerCIS, a tool that we find invaluable as speech scientists. They would also like to thank Althea Bauernschmidt and Lawrence Phillips for their assistance in data collection and analysis, and Luis Hernandez for providing technical assistance and advice in the design and implementation of the experimental procedures. This work supported by NIH-NIDCD Training Grant No. T32-DC00012 and NIH-NIDCD Research Grant No. R01-DC00111.

References

Bradlow AR, Toretta GM, Pisoni DB. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. Speech Commun. 1996; 20:255–272. [PubMed: 21461127]

Burkholder, RA. Research on Spoken Language Processing Technical Report No. Vol. 13. Bloomington, IN: Speech Research Laboratory, Indiana University; 2005. Perceptual learning of speech processed through an acoustic simulation of a cochlear implant.

Chiu C-YP. Specificity of auditory implicit and explicit memory: Is perceptual priming for environmental sounds exemplar specific? Mem. Cognit. 2000; 28:1126–1139.

Chiu C-YP, Schacter DL. Auditory priming for nonverbal information: Implicit and explicit memory for environmental sounds. Conscious Cogn. 1995; 4:440–458. [PubMed: 8750418]

Clark, GM. Learning to understand speech with the cochlear implant. In: Fahle, M.; Poggio, T., editors. Perceptual Learning. Cambridge, MA: MIT Press; 2002. p. 147-160.

Clopper CG, Pisoni DB, Tierney AT. Effects of open-set and closed-set task demands on spoken word recognition. J. Am. Acad. Audiol. 2006; 17:331–349. [PubMed: 16796300]

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. J. Exp. Psychol. 2005; 134:222–241.

Dorman M, Loizou P. The identification of consonants and vowels by cochlear implants patients using a 6-channel CIS processor and by normal hearing listeners using simulations of processors with two to nine channels. Ear Hear. 1998; 19:162–166. [PubMed: 9562538]

Dorman MF, Loizou PC, Rainey D. Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. J. Acoust. Soc. Am. 1997; 102:2993–2996. [PubMed: 9373986]

Egan JP. Articulation testing methods. Laryngoscope. 1948; 58:955–991. [PubMed: 18887435]

Fu Q-J, Galvin J, Wang X, Nogaki G. Moderate auditory training can improve speech performance of adult cochlear implant patients. ARLO. 2006; 6:106–111.

Gaver WW. What in the world do we hear?: An ecological approach to auditory event perception. Ecological Psychol. 1993; 5:1–29.

Gonzales J, Oliver JC. Gender and speaker identification as a function of the number of channels in spectrally reduced speech. J. Acoust. Soc. Am. 2005; 118:461–470. [PubMed: 16119365]

Gygi B, Kidd RR, Watson CS. Spectral-temporal factors in the identification of environmental sounds. J. Acoust. Soc. Am. 2004; 115:1252–1265. [PubMed: 15058346]

Herman, R.; Pisoni, DB. Research on Spoken Language Processing Progress Report No. Vol. 24. Bloomington, IN: Speech Research Laboratory, Indiana University; 2000. Perception of ‘elliptical speech’ by an adult hearing-impaired listener with a cochlear implant: some preliminary findings on coarse-coding in speech perception; p. 87-112.

House AS, Williams CE, Hecker MHL, Kryter KD. Articulation-testing methods: Consonantal differentiation with a closed-response set. J. Acoust. Soc. Am. 1965; 37:158–166. [PubMed: 14265103]

IEEE. IEEE recommended practice for speech quality measurements; IEEE Report No. 297; 1969.

Karl, JR.; Pisoni, DB. Research on Spoken Language Processing Progress Report No. 19. Bloomington, IN: Speech Research Laboratory, Indiana University; 1994. Effects of stimulus variability on recall of spoken sentences: A first report; p. 145-193.

Kidd GR, Watson CS, Gygi B. Individual differences in auditory abilities. J. Acoust. Soc. Am. 2007; 122:418–435. [PubMed: 17614500]

Lachs, L.; McMichael, K.; Pisoni, DB. Speech perception and implicit memory: Evidence for detailed episodic encoding of phonetic events. In: Bowers, J.; Marsolek, C., editors. Rethinking Implicit Memory. Oxford: Oxford University Press; 2003. p. 215-235.

Lewis JW, Wightman FL, Brefczynski JA, Phinney RE, Binder JR, DeYoe EA. Human brain regions involved in recognizing environmental stimuli. Cereb. Cortex. 2004; 14:1008–1021. [PubMed: 15166097]

Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. Ear Hear. 1998; 19:1–36. [PubMed: 9504270]

Marcell MM, Borella D, Greene M, Kerr E, Rogers S. Confrontation naming of environmental sounds. J. Clin. Exp. Neuropsychol. 2000; 22:830–864. [PubMed: 11320440]

Morris CD, Bransford JD, Franks JJ. Levels of processing versus transfer appropriate processing. J. Verbal Learn. Verbal Behav. 1977; 16:519–533.

National Institutes of Health (NIH). Cochlear implants in adults and children; NIH Consens Statement. 1995. p. 1-29.

Obleser J, Wise RJS, Dresner MA, Scott SK. Functional integration across brain regions improves speech perception under adverse listening conditions. J. Neurosci. 2007; 27:2283–2289. [PubMed: 17329425]

Reed CM, Delhorne LA. Reception of environmental sounds through cochlear implants. Ear Hear. 2005; 26:48–61. [PubMed: 15692304]

Shafiro, V. Unpublished doctoral dissertation. New York: CUNY; 2004. Perceiving the sources of environmental sounds with a varying number of spectral channels.

Shannon RV. Speech and music have different requirements for spectral resolution. Int. Rev. Neurobiol. 2005; 70:121–134. [PubMed: 16472633]

Shannon RV, Fu Q-J, Galvin J. The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. Acta Oto-Laryngo. 2004; 552 Suppl.:1–5.

Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science. 1995; 270:303–304. [PubMed: 7569981]

Surprenant AM, Watson CS. Individual differences in the processing of speech and nonspeech sounds by normal hearing listeners. J. Acoust. Soc. Am. 2001; 110:2085–2095. [PubMed: 11681386]

Tice, R.; Carrell, T. Level 16 V2.0.3. Lincoln, NE: University of Nebraska; 1998.

Tulving E, Thomson DM. Encoding specificity and retrieval processes in episodic memory. Psychol. Rev. 1973; 80:352–373.

Tye-Murray N, Tyler R, Woodward G, Gantz B. Performance over time with a Nucleus and Ineraid cochlear implant. Ear Hear. 1992; 13:200–209. [PubMed: 1397761]

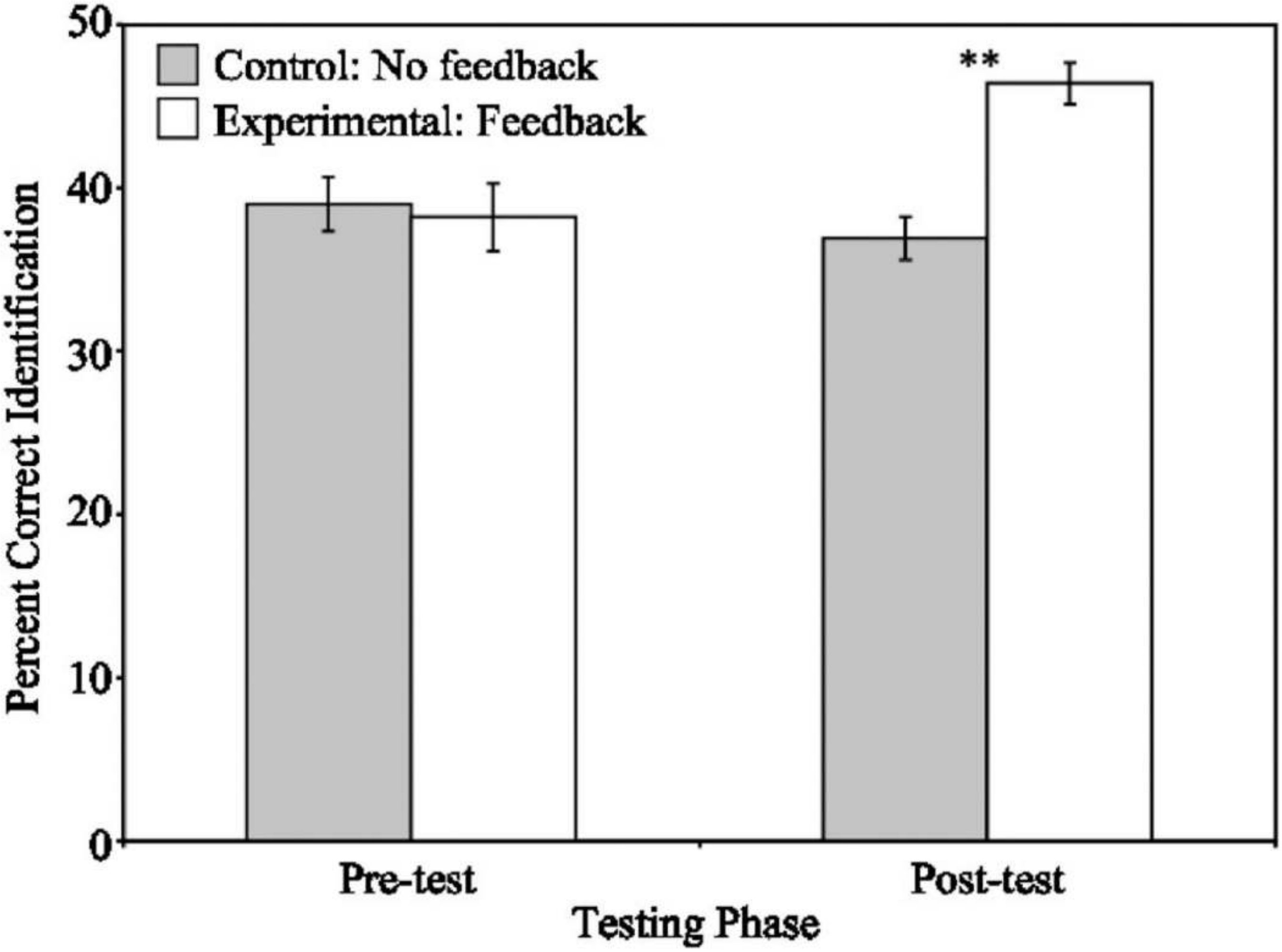

FIG. 1.

Bar graph displaying perceptual accuracy scores at identifying environmental sounds across the experimental and control groups. The type of training that a subject received is indicated on the *x* axis and coded by colored bars (control: gray; experimental: white). Posttest scores only contain the responses to stimuli on which subjects did not receive explicit training (see the text). Asterisks indicate statistically significant differences between groups on pre- or posttest performance ($p < 0.01$).

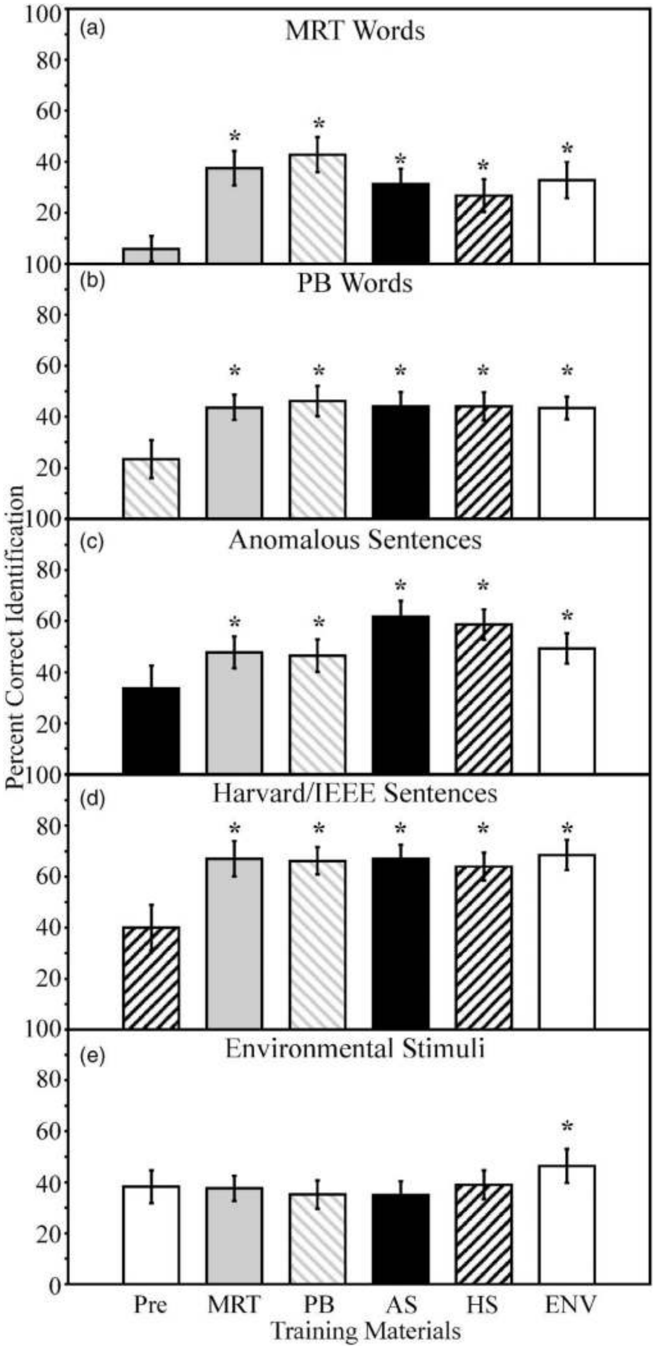

FIG. 2.

Bar graph displaying perceptual accuracy scores at identifying MRT words (A), PB words (B), anomalous sentences (C), Harvard/IEEE sentences (D), and environmental sounds (E) as a function of training. The type of training that a subject received is indicated on the *x* axis and coded by colored bars (MRT: gray; PB: gray striped; AS: black; HS: black striped; ENV: white). Pretest scores correspond to the baseline for each type of stimuli. Posttest scores only contain the responses to stimuli on which subjects did not receive explicit training (see the text). Asterisks indicate statistically significant differences from the baseline for each group ($p < 0.05$).

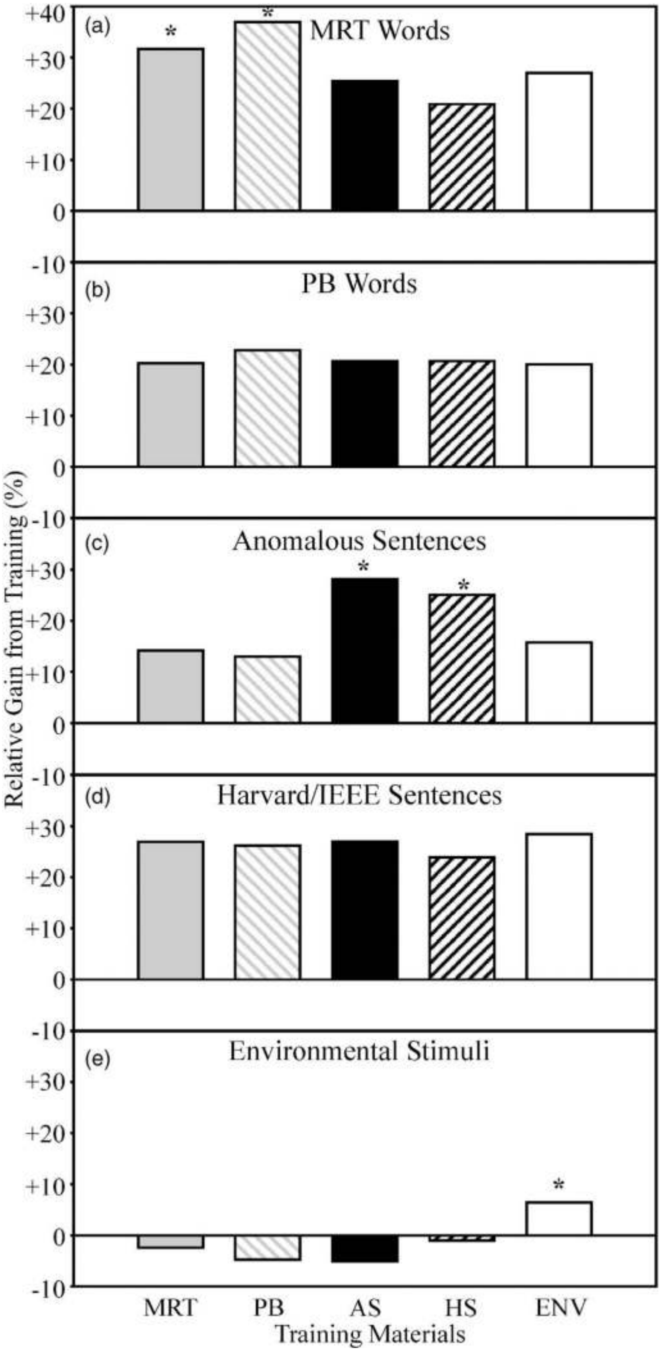

FIG. 3.

Bar graph displaying the relative gains from training for the MRT words (A), PB words (B), anomalous sentences (C), Harvard/IEEE sentences (D), and environmental sounds (E). The type of training that a subject received is indicated along the *x* axis and coded by color (MRT: gray; PB: gray striped; AS: black; HS: black striped; ENV: white). Gain scores were computed by subtracting the posttest or generalization scores from the scores at pretest. Asterisks indicate statistically significant differences from the other groups ($p < 0.05$).

TABLE I

Block design of the experiment identifying the materials presented during each phase. Although all subjects were presented with the same materials, the order in which they were presented varied by block according to the training condition to which subjects were assigned (MRT: modified rhyme test words; PB: phonetically balanced words; HS: Harvard/IEEE sentences: AS: anomalous sentences; and ENV: environmental sounds).

Training	Block 1 Pretest	Block 2 Training	Block 3 Gen. 1	Block 4 Posttest	Block 5 Gen. 2	Block 6 Gen. 3	Block 7 Gen. 4
MRT	MRT	MRT	AS	MRT	HS	PM	ENV
PB	PB	PB	AS	PB	HS	MRT	ENV
AS	AS	AS	MRT	AS	HS	PB	ENV
HS	HS	HS	MRT	HS	AS	PB	ENV
ENV	ENV	ENV	AS	ENV	MRT	HS	PB

TABLE II

Performance at pretest and at posttest across training groups. The individual lists used in the posttest are decomposed into materials from the pretest list, training list and novel list. The final posttest score does not contain the scores from the training list to avoid a confound with feedback.

Training	Pretest	Pretest list	Training list	Novel list	Posttest
MRT	M = 0.06	M = 0.40	M = 0.58	M = 0.35	M = 0.37
	S.D. = 0.06	S.D. = 0.13	S.D. = 0.16	S.D. = 0.12	S.D. = 0.11
PB	M = 0.23	M = 0.44	M = 0.55	M = 0.48	M = 0.46
	S.D. = 0.11	S.D. = 0.09	S.D. = 0.15	S.D. = 0.12	S.D. = 0.09
AS	M = 0.34	M = 0.62	M = 0.78	M = 0.62	M = 0.62
	S.D. = 0.14	S.D. = 0.12	S.D. = 0.13	S.D. = 0.13	S.D. = 0.10
HS	M = 0.40	M = 0.72	M = 0.97	M = 0.56	M = 0.64
	S.D. = 0.20	S.D. = 0.10	S.D. = 0.07	S.D. = 0.12	S.D. = 0.08
ENV	M = 0.38	M = 0.40	M = 0.50	M = 0.53	M = 0.46
	S.D. = 0.10	S.D. = 0.14	S.D. = 0.11	S.D. = 0.11	S.D. = 0.11

TABLE III

Correlations between the various stimuli presented at posttest and during generalization for each training group (see Table I for abbreviations). Rows are blocked by training condition and the specific materials used during posttest are indicated by italicized font (posttest values only contain the subset of materials on which subjects did not receive explicit feedback). Values along the diagonal indicate the percent correct recognition scores for the stimuli.

Training		MRT	PB	AS	HS	ENV
MRT	*MRT*	37%	$r=0.77^{**}$	$r=0.57^{*}$	$r=0.67^{**}$	n.s.
	PB	···	43%	$r=0.55^{*}$	$r=0.65^{**}$	n.s.
	AS	···	···	48%	$r=0.90^{**}$	n.s.
	HS	···	···	···	67%	n.s.
	ENV	···	···	···	···	38%
PB	MRT	43%	$r=0.66^{***}$	$r=0.51^{**}$	n.s.	$r=0.55^{**}$
	PB	···	46%	$r=0.53^{**}$	n.s.	n.s.
	AS	···	···	48%	n.s.	n.s.
	HS	···	···	···	66%	n.s.
	ENV	···	···	···	···	35%
AS	MRT	31%	n.s.	n.s.	n.s.	n.s.
	PB	···	44%	n.s.	n.s.	$r=0.45^{*}$
	AS	···	···	62%	$r=0.63^{***}$	n.s.
	HS	···	···	···	67%	n.s.
	ENV	···	···	···	···	35%
HS	MRT	27%	$r=0.59^{**}$	$r=0.62^{***}$	$r=0.40^{*}$	n.s.
	PB	···	44%	$r=0.73^{***}$	$r=0.51^{**}$	$r=0.57^{**}$
	AS	···	···	59%	$r=0.55^{**}$	n.s.
	HS	···	···	···	64%	n.s.
	ENV	···	···	···	···	39%
ENV	MRT	33%	n.s.	n.s.	n.s.	n.s.
	PB	···	43%	n.s.	n.s.	n.s.
	AS	···	···	49%	$r=0.69^{***}$	n.s.
	HS	···	···	···	68%	n.s.
	ENV	···	···	···	···	46%

Only statistically significant correlations are displayed (* $p < 0.05$;

** $p < 0.01$,

*** $p < 0.001$).