

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 28 (2007)
Indiana University

Perceptual Learning Under a Cochlear Implant Simulation¹

Jeremy L. Loebach and David B. Pisoni

*Speech Research Laboratory
Department of Psychological and Brain Sciences
Indiana University
Bloomington, Indiana 47405*

¹ This research supported by NIH NIDCD R01 Research Grant DC00111, and NIH NIDCD T32 Training Grant DC00012 to Indiana University. I wish to thank Althea Bauernschmidt for her assistance in data collection and analysis, without her I would have had to work so much harder! I would also like to thank Luis Hernandez for providing technical assistance and advice in the design and implementation of the experimental procedures.

Perceptual Learning Under a Cochlear Implant Simulation

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/post-test design to evaluate the efficacy and generalization of training in normal hearing subjects listening to CI simulations (8-channel sinewave vocoder). Subjects were trained with words (simple or complex), sentences (meaningful or anomalous), or environmental stimuli, and then were tested using an open-set identification task. Subjects were trained on only one type of material but were tested on all materials. All groups showed significant improvement as a result of 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 signals; however, training on environmental signals successfully generalized to speech. These data demonstrate that the perceptual learning of degraded speech is highly context-dependent and that the specific stimulus materials that a subject experiences during training have a substantial impact on generalization to new materials.

Introduction

Despite the recent advances in cochlear implant (CI) technology, a large amount of variability in outcome and benefit is consistently reported among CI users. Although differences in etiology, onset, and duration of deafness, age at implantation and physiological factors (electrode insertion depth, availability of viable neurons, etc.) can account for a portion of this variability (NIH, 1995), a considerable amount of variability remains unexplained. The absence of rigorous standardized training regimens confounds the issue at a fundamental level. The experiences of CI users may differ from the start, leading to differences in auditory perceptual learning during adaptation to their prostheses. Could the standardization of training establish a more stable foundation, and allow the dissociation of audiological factors from other neural and cognitive factors? Moreover, what type of training is most effective, and yields the most robust levels of generalization to new materials? The present study seeks to investigate the efficacy of different training regimens in normal hearing subjects using both speech and environmental stimuli that have been 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 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 signals. While clinicians often cite this benefit as part of the rationale for implantation, the degree to which CI users can actually recognize and identify environmental signals is largely unknown (cf., Reed & Delhorne, 2005).

Research using acoustic simulations of CIs has met with great success. From the earliest simulations of Shannon and colleagues (Shannon, Zeng, Kamath, Wygonski & Ekelid, 1995), the effectiveness and utility of acoustic models of CIs has been apparent. The vocoder model of a CI simulates the limited number of spectral channels available in the electrode array by dividing the acoustic

signal using a series of band-pass filters. Band limited noise replaces the spectrum of each band to simulate the effect of wide-band electrical stimulation of each electrode. The amplitude envelope, which is derived from the original bands using a low pass filter, is then used to modulate the noise to simulate the temporal profile of electrical stimulation at each electrode. The result is a signal that acoustically simulates the spectrally degraded conditions that CI users may normally encounter.

The seminal work of Shannon and colleagues using the vocoder demonstrated that high levels of speech recognition persisted despite such radical spectral degradation (Shannon et al., 1995). Using signals with one, two, three or four spectral channels, Shannon demonstrated that when more spectral channels were available to the listeners, higher levels of perceptual identification were observed. A single channel provided sufficient information to allow moderately accurate closed-set recognition of English consonants and vowels (48% correct and 35% correct respectively), and as the number of channels increased so did recognition rates. For vowels and meaningful sentences, asymptotic performance (> 90% correct) was reached with three channels, whereas recognition rates for consonants continued to increase from three to four channels. Consonant recognition was far less robust than vowel recognition due to several factors. When consonants were classified according to the guidelines of Miller and Nicely (1955), perception of manner and voicing cues reached asymptote with just 2 spectral channels (> 90% correct identification), as compared to classification based on place of articulation, which never exceeded 60% correct even with four channels. These data demonstrate the robust nature of the human perceptual system, which can perform well even when spectral information is severely limited, so long as sufficient temporal information is preserved (Shannon et al., 1995).

Follow-up studies further refined the methodology, expanding the number of channels available and altering the carrier used. When the maximum number of spectral channels available was incrementally increased from 4 to 9, asymptotic performance was observed for closed-set vowels with 8 channels, and meaningful sentences with 5 channels (Dorman, Loizou & Rainey, 1997). Consonant identification reached asymptote with 6 channels, which was a result of increased accuracy in identifying place of articulation, which also reached asymptote at 6 channels (Dorman et al., 1997). Moreover, the type of carrier used did not appear to have an adverse effect on performance. In their original study, Shannon and colleagues used a noise vocoder, in which white noise was used to remove the spectral detail from each band (Shannon et al., 1995). The anecdotal reports of CI users, however, were not of hearing bursts of noise, but of hearing “beep tones” (Dorman et al., 1997), raising the question of whether noise is the most appropriate carrier to use (Dorman et al., 1997). Using a sinewave vocoder, which replaces the spectral detail of each band with a sinusoid anchored at the band center, Dorman and colleagues demonstrated that performance did not differ from that observed using the noise vocoder (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 stimuli, demonstrating that the vocoder can successfully simulate the output of a CI in order to elicit equivalent levels of performance (Dorman & Loizou, 1998).

Although studies using the noise and sinewave vocoders have focused primarily on the identification of linguistic content (e.g., isolated consonants and vowels), the real world is composed of many other complex auditory events that are transmitted via the acoustic signal. Compared to speech, considerably less is known about the perception of environmental sounds, both in the clear and processed by vocoder models. Environmental signals are very useful for neuropsychological and cognitive evaluation because they can assess basic sensory and cognitive capabilities without the added dimensions of linguistic information and context. Although there may be some commonalities between the perceptual systems required for the identification of speech and environmental stimuli, the degree to which they operate independently is unknown. Some cross-modal priming has been observed for environmental

stimuli. When the acoustic presentation of an unprocessed environmental stimulus is paired with the orthographic presentation of the stimulus name during the study phase of an experiment, subjects are faster and more accurate at identifying the stimulus during the test phase as compared to if they saw the name presented without the sound (Chiu & Schacter, 1995). This priming effect is context specific, however. If the exemplar is sufficiently different from the test stimulus, the strength of priming is reduced (Chiu, 2000). For example, if the subject received one exemplar of an environmental stimulus during the study phase (such as the sound of a bird chirping), but was tested with a different exemplar (a different bird chirping), priming was significantly reduced. In speech, the stimulus-specific form can also be preserved in addition to the more abstract symbolic lexical form (Lachs, McMichael & Pisoni, 2003). Thus, at least at a surface level, it appears that environmental stimuli may be encoded in a similar manner to speech.

Although the processing of environmental signals may share some similarity with the neural and cognitive processes used to perceive speech, many of the acoustic (spectral and temporal) characteristics of speech are fundamentally different from environmental signals (see Stevens, 1980 for example). In a series of recent experiments investigating the perceptual identification of environmental signals, Gygi and colleagues trained subjects to identify 70 environmental stimuli in the clear using a three-letter code (Gygi, Kidd & Watson, 2004). They then processed the stimuli using a series of low, high, and band-pass filters and tested subjects over a period of nine days. Overall, they found that both speech and environmental stimuli may share a similar range of critical frequencies that are important for identification. The most important acoustic information for the recognition of environmental stimuli lies between 1200 and 2400 Hz, which is identical to the region identified as crucial for speech under the Articulation Index (Gygi et al., 2004). Moreover, even when the stimuli were low-pass or high-pass filtered at the extremes (300 and 8000 Hz), recognition remained higher than 50% correct (Gygi et al., 2004).

When the stimuli were processed using one and six-channel noise vocoders, the results were more variable. Naïve subjects in both groups showed significant improvement over a two-day period (1-channel: 13% correct on day 1 to 23% correct on day 2; 6-channel: 36% correct on day 1 to 66% correct on day 2), but performance was significantly higher for the 6-channel stimuli (Gygi et al., 2004). Not surprisingly, the stimuli that showed the greatest improvement were those that had broader harmonic structure and spectral detail (Gygi et al., 2004). However, these results should be considered with some caution because certain aspects of performance may be attributable to task familiarity. A group of subjects who were first trained to criterion on the unprocessed stimuli performed significantly better on the 1-channel stimuli than did the naïve subjects (Gygi et al., 2004), which could be attributable to increased experience with the three letter codes rather than to familiarity with the stimuli themselves. In addition, Gygi and colleagues used a closed set recognition task, which constrains the possible choices that subjects can make to those within a specified stimulus set. Subjects could have been systematically eliminating the possible alternatives as they became increasingly familiar with the test set.

Using a slightly different task, Shafiro demonstrated that the reliance on spectral and temporal information in the recognition of environmental stimuli processed with a noise vocoder may be different than is observed for speech (Shafiro, 2004). Sixty environmental stimuli were processed with 2, 4, 8, 16 and 32 channel vocoders, and presented to normal hearing subjects using a Latin square design, such that each subject only heard one version of a stimulus, but all band conditions were presented across all subjects. In general, improved closed set recognition (out of 60) was observed as the number of channels increased. With only 2 channels, performance was low (32% correct), but reached asymptote at 66% correct with 16 channels (Shafiro, 2004). Moreover, the performance depended on the stimulus itself: while some environmental stimuli showed increases in accuracy with the addition of more spectral

channels, others showed decreases (Shafiro, 2004). In particular, stimuli that relied more on spectral information (e.g., church bell, birds chirping) showed increases, whereas those that relied more on temporal information (e.g., clapping, footsteps) showed decreases. Thus, it appears that some environmental stimuli may show an altogether different pattern of spectro-temporal dependence as compared to speech signals.

Relatively few studies have examined the perception of environmental stimuli by CI users. Using a closed-set testing format, Tye-Murray and colleagues assessed the abilities of fourteen CI users to identify 36 environmental stimuli at 1, 9, 18 and 30 months post-implantation (Tye-Murray, Tyler, Woodward & Gantz, 1992). Overall, performance increased significantly over time, from about 32% correct at 1 month, to 38% correct at 9 months, and topping out at 42% correct at 18 months (Tye-Murray et al., 1992). These gradual changes were statistically significant, although far slower than the gains typically observed for speech. A more recent study by Reed and Delhorne (2005) compared environmental sound recognition and NU-6 word identification. Environmental stimuli were organized into four thematic lists of ten stimuli each, and subjects made closed set responses by clicking one of ten buttons presented on a computer screen. Performance of the eleven CI users differed across the four lists of environmental stimuli, with a mean identification score of 79% correct (Reed & Delhorne, 2005). Average performance on the closed set environmental stimulus identification was significantly better than performance on the open set word identification, which was only 39% correct (Reed & Delhorne, 2005). Subjects were divided into high performing and low performing groups based on the median score for word identification (34% correct). High performing subjects (> 34%) performed better at identifying environmental stimuli than did low performing subjects (Reed & Delhorne, 2005). The authors hypothesized that the differences in performance may be due to differences in exposure to environmental stimuli in their daily environment (Reed & Delhorne, 2005). However, it is unclear whether additional exposure or standardized training could increase the performance of the low performing subjects.

One common theme throughout the studies using vocoded signals is the issue of perceptual learning. Even though subjects can accurately identify speech processed by a vocoder, a period of adjustment is frequently required. In the original Shannon study, subjects received 8-10 hours of exposure to the synthesis condition in order to adapt to the stimuli and stabilize their performance (Shannon et al., 1995). Explicit training on the testing materials was used in the studies by Dorman and colleagues in order for subjects to “warm up” to the synthesis condition (Dorman et al., 1997; Dorman & Loizou, 1998). Although some type of auditory training is necessary when adapting to acoustic simulations of CIs, the best and most efficient form that maximizes perceptual learning and promotes robust generalization and transfer to other materials has not been adequately examined.

In a series of recent experiments, Davis and colleagues investigated the use of lexical information during adaptation to 6-channel noise vocoded sentences (Davis, Johnsrude, Hervais-Adelman, Taylor & McGettigan, 2005). Five experiments were conducted in order to examine the mechanisms of perceptual learning. In the first experiment, they assessed whether exposure to the stimulus materials without any feedback results in perceptual learning. Subjects were presented with a set of thirty sentences that were processed with the vocoder and asked to transcribe as much of each as possible. Open set identification increased significantly across the 30 sentences, from 32% correct keyword identification on the first ten sentences to 43% correct on the last 10 sentences. These gains can be attributed to perceptual attunement to vocoded speech, since subjects received no feedback.

The effectiveness of auditory feedback was assessed in Experiment 2. Like Experiment 1, subjects transcribed each sentence; however, after they made their response they were provided with auditory feedback. One group heard the “distorted” sentence followed by the unprocessed version

(DDC), whereas the other group heard the clear sentence followed by the repetition of the distorted version (DCD) in order to elicit stimulus pop out. Both groups showed significant gains as a result of training, increasing from 43% to 73% correct from the first to final 10 sentences for subjects in the DDC group, and from 50% to 77% correct for subjects in the DCD group. Although both groups showed equivalent gains, the group who received the DCD training experienced performed significantly better (Davis et al., 2005).

The typical CI user will not have access to the unprocessed version of a stimulus, however, so in Experiment 3, Davis and colleagues explored whether the addition of the orthographic version of the sentence enhances perceptual learning (Davis et al., 2005). Subjects either received feedback in the form of the repetition of the distorted version of the sentence paired with and without the written transcription. Subjects who were presented with the repetition of the distorted sentence alone showed significant improvement, increasing from 38% correct to 69% correct. Subjects who also received the orthographic form of the stimulus performed significantly better, improving from 50% to 77% correct, a level of performance identical to those subjects in experiment 2 who experienced stimulus pop out. Such comparable improvement suggests that presentation of the orthographic form of the sentence is just as effective as presentation of the original unprocessed acoustic version (Davis et al., 2005).

Although these gains are impressive, one potential factor that could contribute to the results of the first three experiments is the use of contextually constrained meaningful sentences. When Davis and colleagues controlled the amount of lexical information in the sentences, however, the amount of learning varied (Davis et al., 2005). Subjects who were trained with sentences comprised entirely of non-words improved with training, but performed significantly more poorly than those who were trained on meaningful sentences (Experiment 4). This experimental task may be more difficult, however, given that the materials are not valid English words. To examine these effects more in more detail, Davis and colleagues conducted a final experiment that systematically varied the amount of lexical information. Subjects were trained on meaningful sentences, semantically anomalous sentences (sentences where the function words are correctly placed, but the content words are unrelated), non-word sentences, or Jabberwocky sentences (anomalous sentences where the content words are replaced by non-words). All groups showed improvement over the training interval, and two distinct groups emerged based on performance. Subjects 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 may be required in order to elicit effective levels of learning (Davis et al., 2005).

The results reported by Davis and colleagues raise several important questions. Although they demonstrated that feedback significantly influences performance, the type of feedback they used would not necessarily apply to the typical CI user. In an individual with electric hearing, there is never an opportunity for the presentation or repetition of the unprocessed 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 clear version suggests that such feedback could be useful to CI users. In addition, subjects who did not receive explicit feedback showed significantly lower levels of performance overall, but still showed similar gains due to training.

In a more comprehensive study, Burkholder and colleagues (Burkholder, 2005; Burkholder, Svirsky & Pisoni, submitted 1; 2) 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 post-test gains than receiving the unprocessed version alone. Moreover, subjects who were trained on the anomalous sentences showed identical pre to post-test gains as subjects

trained on meaningful sentences, but showed significantly greater benefits during generalization to new materials including environmental stimuli (Burkholder, 2005; Burkholder et al., submitted 1; 2). 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. This point is underscored in the observation that training on speech stimuli successfully generalized to the identification of environmental stimuli.

One limitation of the studies by Burkholder and colleagues is that they only assessed the generalization of training with speech to environmental stimuli, but not the converse. If subjects are relying on the acoustic structure of the stimuli, one would predict that training on environmental stimuli should successfully generalize to speech, an issue that we address in the current work. In addition, no baseline identification data were collected for the environmental stimuli, so it is unknown if the subjects were performing significantly better at identifying the environmental stimuli 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 vice versa. In addition, previous studies have shown that training on simple CV and CVCs may produce only modest gains in performance on sentence identification (Fu, Galvin, Wang & Nogaki, 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?

As there are currently no standard rehabilitation protocols following cochlear implantation, understanding how the perceptual learning of spectrally degraded stimuli transfers to new materials is especially relevant. Evaluating the strengths and weaknesses of different training paradigms is critical for the development of rehabilitation strategies that maximize perceptual learning and promote robust generalization to new materials. The purpose of the present study, therefore, was to examine the effect of training on the recognition of speech and environmental stimuli processed by a sinewave vocoder. Specifically, we assessed the perceptual learning of CVCs, words, meaningful sentences, anomalous sentences and complex non-speech environmental stimuli using a pre/post-test design, and compared the generalization to different materials.

Method

Subjects

One hundred thirty normal-hearing adults from the IU community participated in the study. Of the 130 subjects, 95 were female, 34 were male, and one self reported being transgender. Subjects ranged in age between 18 and 60, with a mean age of 22.7 years. All subjects reported having uncorrected normal hearing and that English was the first language that they learned in infancy. Most subjects (n= 117) were monolingual; although a small number reported being fluent bi- (n= 11) or tri-lingually (n= 2). Subjects were given credit in their Introductory Psychology course for their participation (n= 34), or were paid at the rate of \$10 per hour (n= 96).

Of the 130 subjects, five were excluded from the final data analysis. One subject was excluded after reporting that he/she could not hear the stimuli as speech. One subject was excluded 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. Two 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, and the other typed only gibberish (random keystrokes) rather than making a meaningful response to the stimuli.

Stimuli

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

Modified Rhyme Test. The Modified Rhyme Test (MRT) corpus consisted of 300 words organized into fifty lists, where each list contains six rhymed variations on a common syllable (House, Williams, Hecker & Kryter, 1965). Within each list, the word initial or word final consonant is systematically varied to produce six items each differing only by a minimal pair (e.g., “bat”, “bad”, “back”, “bass”, “ban”, “bath”). Stimuli consisted of ninety 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. A female talker produced forty-two of the words, and a male talker, the remaining forty-eight.

Phonetically Balanced Words. The Phonetically Balanced corpus (PB) consisted of twenty lists of fifty monosyllabic words whose phonemic composition approximates the statistical occurrence in American English (e.g., “bought”, “cloud”, “wish”, “scythe”) (Egan, 1948). Stimuli consisted of ninety 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.

Harvard/IEEE Sentences. The Harvard/IEEE Sentence database consisted of seventy-two lists of ten 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 twenty-five sentences drawn from lists 1-10 of the Harvard/IEEE Sentence database and their associated wav file recordings that were obtained from the speech corpus originally created by Karl and Pisoni (1994). A female talker produced fourteen sentences and a male talker produced the remaining eleven. 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, Toretta & Pisoni, 1996).

Anomalous Harvard/IEEE Sentences. Semantically anomalous sentences preserve the canonical syntactic structure of English, but have no meaning. The anomalous sentences from the corpus of Herman and Pisoni (2000) used the Harvard/IEEE sentence materials to create phonetically balanced meaningless sentences. The keywords from the 100 sentences in lists 11-20 were coded according to semantic category (noun, verb, adjective, adverb) and replaced with words from equivalent semantic categories from lists 21-70 (Herman & 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 semantic context to identify the keywords. Stimuli consisted of twenty-five anomalous sentences drawn from the anomalous Harvard/IEEE sentences corpus of Herman and Pisoni (Herman & 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.

Environmental Stimuli. The environmental signal database of Marcell and colleagues consists of stimuli recorded from a wide variety of acoustic environments developed for use in neuropsychological evaluation and confrontation naming studies (Marcell, Bordella, Greene, Kerr & Rogers, 2000). The 120 stimuli in the corpus contain 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), non-speech sounds produced by humans (snoring, crying, coughing), musical instruments (piano, trumpet, flute), tools (hammer, vehicles), liquids (water boiling, rain) among others. These signals have been normed in a group of neurologically intact 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, Bordella, Greene, Kerr & Rogers, 2000). Stimuli consisted of ninety environmental stimuli and their associated wav file recordings obtained from a digital database published by the authors on the Internet (<http://www.cofc.edu/~marcellm/confront.htm>). Stimulus selection from a variety of acoustic categories provided a wide representation of sound types and familiarity ratings.

Synthesis

Stimulus processing used a freeware program (Tiger CIS) developed for research that is available on the Internet (<http://www.tigerspeech.com/>). The software simulated an 8-channel CI using the CIS processing strategy. Stimulus processing involved two phases, an analysis phase, which divides the signal into 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 8 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 & Carrell, 1998) ensured that stimuli were equal in intensity across all materials, and that no peak clipping occurred.

Materials

Data collection used a custom script written for PsyScript, and implemented on four Apple PowerMac G4 (512 Mb RAM) computers running OS 9.2.2, and four 15 inch color Sony LCD monitors (1024x768 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 dBv 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.

Procedures

All methods and materials were approved by the Human Subjects Committee and Institutional Review Board at Indiana University Bloomington. Informed consent was established before beginning the experiment, and subjects were given a short subject information form asking for basic background information (basic background, demographic and contact information) and inquiring as to any prior hearing, speech, or language problems.

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. 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, presented at the center of the screen for 500 milliseconds alerted the subject as to the upcoming trial. The fixation cross was erased, and the sound file was presented at the next vertical retrace. 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 allowed to rest between each trial as needed. The experimental session lasted on average 45 minutes. All subjects received written and verbal debriefing after the experiment.

Training

Each training condition consisted of seven blocks. Stimuli were pre-randomized, 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 (Fig. 1). Each condition began with a pre-test (block 1), which assessed the subjects’ ability to identify the materials before training began. At this point, subjects are naïve to the stimulus processing, and had received no familiarization or adaptation. During the training sessions, subjects heard a stimulus, and then responded in the dialog box that appeared on the computer screen. Following their response, 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 (the transcription of the word or sentence, or the descriptive label of the environmental stimuli) irrespective of whether their previous response was correct. An intervening generalization block occurred between the training block (block 2) and the post-test block (block 4). During the post-test, subjects heard a selection of old materials from the pre-test and post-test, as well as new materials from the same category. The post-test materials were selected to assess the effects of explicit training (using training materials), familiarity without explicit training (pre-test materials) and novelty (previously unheard materials). The remaining three blocks were generalization blocks testing the effects of training.

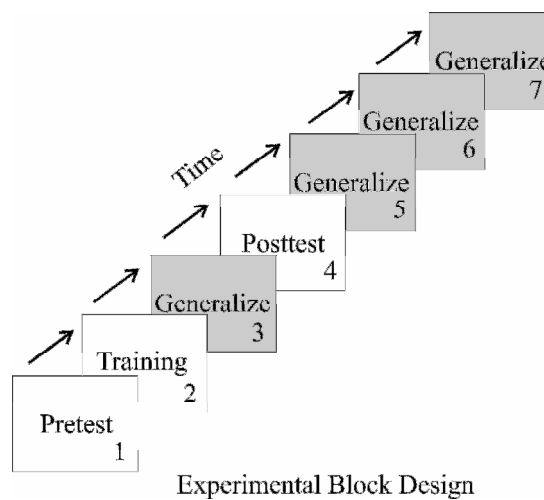


FIGURE 1. Block design of the experimental trials for all training groups.

MRT Word Training. During the pre-test, listeners were presented with twenty MRT words. Training consisted of fifty novel MRT words. An intervening generalization block occurred in block 3 to prevent habituation to the stimuli, and consisted of twenty-five anomalous sentences. The post-test in block 4 presented a total of 60 MRT words, twenty of which were drawn from the pre-test materials, twenty from training, and twenty were novel stimuli with which subjects had no previous experience in the experiment. The remaining three blocks consisted of generalization to 25 meaningful sentences (block 5), 50 novel PB words (block 6) and 60 environmental signals (block 7).

PB Word Training. PB training utilized an identical design to the MRT training, except that PB words consisted of the pre-test, training, and post-test materials and block 6 consisted of generalization to 50 novel MRT words.

Harvard/IEEE Sentence Training. In order to balance for the relative effect of words transcribed across sentences, fewer sentences were selected. The pre-test block consisted of four Harvard/IEEE sentences (20 key words); the training block consisted of ten novel Harvard/IEEE sentences (50 key words). Block 3 was an intervening generalization block, consisting of 50 MRT words. The post-test in block 4 utilized 12 Harvard/IEEE sentences, 4 selected from the pre-test, four from the post-test and four novel sentences (60 keywords). The remaining three blocks tested the effects of generalization to new materials. Block 5 consisted of 25 anomalous sentences, block 6 of 50 PB words and block 7 of 60 environmental signals.

Anomalous Sentence Training. Anomalous sentence training utilized an identical design to the Harvard/IEEE sentence training, except that the pre-test, training and post-test materials consisted of Anomalous sentences, and block 6 consisted of generalization to 25 novel Harvard/IEEE sentences.

Environmental Stimulus Training. Like the MRT and PB training, training on environmental training stimuli began with a pre-test consisting of twenty environmental signals and training consisting of fifty novel environmental signals. An intervening generalization block occurred in block 3 in order to prevent habituation to the stimuli and consisted of twenty-five Anomalous sentences. The post-test in block 4 presented a total of 60 environmental signals, twenty of which were drawn from the pre-test materials, twenty from training, and twenty 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).

Analysis and Scoring

A supervised spellchecker corrected the more obvious spelling errors and standardized spelling across subjects by changing homophones into a standard spelling. An automated macro searched for target/response matches using a pre-ordained target list, the result of which was then hand checked by a trained research assistant. Responses that were morphologically related to the target were scored as incorrect. PB and MRT words were scored based on whether the entire word was correct, whereas anomalous and meaningful sentences were scored for keywords correct (5 keywords per sentence).

Environmental stimuli were checked using a similar procedure, except more options were included in the target list given the complexity of the stimuli. Scoring rules were modified slightly from those originally used by Marcell and colleagues (Marcell, Bordella, Greene, Kerr & Rogers, 2000) given the nature of the degradation. Animal and insect sounds 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). Responses were considered incorrect if the subject

failed to disambiguate the perceived agent from multiple agents (e.g., ‘whistling’ was an incorrect response for ‘birds’ given that human ‘whistling’ was a viable target, however ‘tweet’ and ‘chirping’ were considered correct). Failure to specify agent, or incorrectly specifying agent was scored as an incorrect response (e.g., for ‘seal’ the response ‘seal barking’ is correct, but the response ‘barking’ is incorrect given that the agent is not specified and could refer to a dog). 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. Multiple instruments from a given class were considered as viable options so long as they afforded a common action (e.g., the responses ‘viola’ and ‘violin’ were considered correct options for the target ‘violin’, however ‘string’ and ‘guitar’ were incorrect responses given that the action affords the use of a bow, whereas the action afforded by the latter response requires plucking).

Non-speech sounds produced by humans were considered correct if they correctly identified the sound given that the agent was unambiguous (e.g., ‘child coughing’ has the possible correct response options of ‘child coughing’, ‘coughing’ or ‘cough’). ‘Scream’ on the other hand was correct if subjects identified the target ‘scream’ or some variant supposing a human agent. ‘Monkey screaming’ was incorrect given the misidentification of the agent. Liquid sounds were considered correct if the subject identified the agent or the action, and allowed for multiple specific sources as appropriate (e.g., ‘water boiling’ had the possible correct options of ‘boil’, ‘bubble’, ‘bubbling’ or ‘bong’).

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 post-test performance. Post-test scores were balanced by only averaging the responses to the materials on which subjects were not explicitly trained, to avoid biasing the findings. The differences in performance on the various post-test materials (items from pre-test, training and novel lists) were assessed with a one-way ANOVA and post hoc Tukey tests. Scores were organized in a column, and coded to reflect the source (pre-test, training or novel). Other paired *t*-tests were conducted to assess the effects of context (Anomalous sentences vs. Harvard/IEEE sentences) and complexity (PB words vs. MRT words). A correlational analysis examined the relationship between performance across blocks to assess whether performance on one type of material was correlated with performance on another. Between subjects comparisons assessed the effects of training on materials across training conditions using one-way Analysis of Variance and post-hoc Tukey tests.

Results

Within Group Comparisons

MRT Training. Overall, initial performance of the 25 subjects who received training on the MRT materials started out very poor, but increased following training (Fig. 2). Percent correct recognition increased from 5.8 % correct at pre-test to 37.5% after training, demonstrating a gain of nearly 32 percentage points. A paired *t*-test indicated that the effect of training was highly significant ($t(1, 24)=13.576, p<0.001$). Comparison of the various post-test materials (data not shown) demonstrated that subjects performed best on stimuli from the training list (materials on which they were explicitly trained), followed by stimuli from the pre-test list (materials with which they were familiarized but not trained) and finally stimuli from the novel list (MRT materials that did not appear before the post-test). A one-way ANOVA revealed a significant main effect of source material, demonstrating that subjects performed differently on materials from the pre-test, training and novel lists ($F(2, 74)=18.967, p<0.001$). Post hoc Tukey tests revealed that subjects performed significantly better on the materials that they heard

during training (58% correct) than on pre-test (40.4% correct) or novel materials (34.6% correct, both $p < 0.001$), demonstrating a significant effect of feedback, and indicating good retention of training. Subject performance did not differ on the materials drawn from the pre-test and novel lists ($p = 0.313$), suggesting that explicit training promotes more of a benefit than exposure alone.

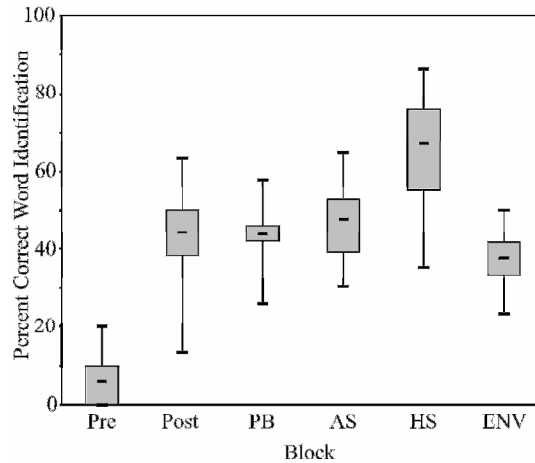


FIGURE 2. Box plot displaying the perceptual accuracy scores as a function of experimental block for the 25 subjects trained to identify the MRT stimuli. Boxes encompass the middle 50% of the data, and horizontal lines indicate the average score for that block. Pre-test scores reflect the baseline performance on the MRT words before training, when subjects were naïve to the processing condition. Post-test scores contain only the responses to MRT stimuli on which subjects did not receive explicit training (see text). MRT and PB words were judged as correct if the subject typed the entire word correctly. Harvard/IEEE (HS) and Anomalous (AS) sentence scores reflect the percent of key words correctly typed. Environmental stimuli (ENV) scores reflect the correct identification of the sound (see text).

Overall, subjects performed best on the Harvard/IEEE sentences (67.0% correct), followed by anomalous sentences (47.7% correct), PB words (43.7% correct) and Environmental stimuli (37.6% correct). A paired t-test revealed a significant effect of sentence context on recognition. Subjects performed significantly better on the Harvard/IEEE sentences than on the anomalous sentences ($t(1,24) = 18.327$, $p < 0.001$). The difference between the scores for the meaningful and anomalous sentences suggests that the addition of context leads to improvement by almost 20%. A paired t-test comparing performance on the MRT and PB words also indicates a difference in performance, with subjects performing significantly better on PB materials than on MRT ($t(1,24) = 3.928$, $p = 0.001$). This may be due to differences in the difficulty of the words used in the MRT and PB lists, since the MRT words include only minimal pairs.

Correlations of the performance across blocks revealed several significant results. Performance at post-test was significantly correlated with performance on each measure except for environmental stimuli (MRTpost-test vs. PB $r = 0.766$, MRTpost-test vs. HS $r = 0.672$, MRTpost-test vs. AS $r = .576$, all $p < 0.01$). Similar relationships were observed for the PB words (PB vs. HS $r = .654$, PB vs. AS $r = .552$), and anomalous and Harvard/IEEE sentences (AS vs. HS $r = .905$). 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 followed by PB and MRT words.

PB Training. Subjects trained on the PB words started out better than those subjects trained on the MRT words. Performance at pre-test was 23.4%, but increased to 46.2% correct following training (Fig. 3). A paired samples t-test indicated that subjects performed significantly better at post-test as compared to pre-test ($t(1,24)=7.134, p<0.001$). Examination of the post-test materials (data not shown) revealed that subjects performed best on stimuli on which they were explicitly trained (55.4% correct), followed by novel PB words (48.2% correct) and words on which they were previously exposed, but not explicitly trained (44.2% correct). A one-way ANOVA revealed a significant main effect of list ($F(2, 74)=5.484, p=0.006$). Post hoc Tukey tests revealed that subjects performed significantly better on materials from the training list ($p=0.005$) than on materials from the pre-test list, but no difference was observed when compared with the materials drawn from the novel list ($p=0.097$). More importantly, subject performance did not differ

As observed for the MRT training condition, subjects performed best on the Harvard/IEEE sentences (66.2% correct), followed by the Anomalous sentences (48.2% correct), MRT words (42.8% correct) and Environmental stimuli (35.2% correct). A paired t-test revealed that subjects performed significantly better on the Harvard/IEEE sentences than on the anomalous sentences ($t(1,24)=12.214, p<0.001$). Subtraction of the scores for the anomalous sentences from those for the Harvard/IEEE sentences reveals a 20% gain from context. Subjects' performance did not differ significantly between the PB words and MRT blocks ($t(1,24)=1.855, p=.076$) although a slight numerical trend was observed favoring PB words.

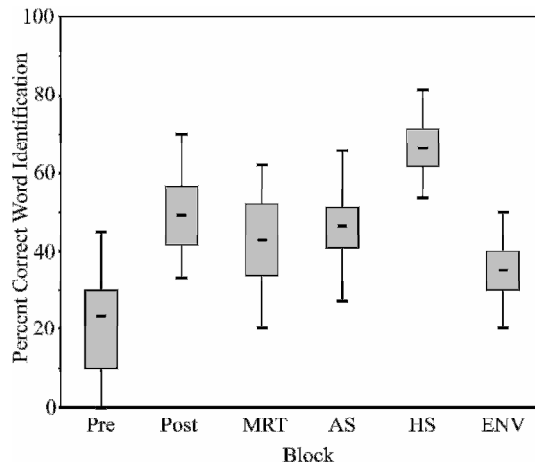


FIGURE 3. Box plots displaying the perceptual accuracy scores as a function of experimental block for the 25 subjects trained to identify the PB stimuli.

Performance on the PB words during the post-test was significantly correlated with performance in the MRT block ($r=.658, p<0.001$), Harvard/IEEE sentences ($r=.539, p=0.005$), but not for the Anomalous sentences or Environmental stimuli. Performance on the Harvard/IEEE sentences was significantly correlated with performance on Anomalous sentences ($r=.609, p=0.001$) and MRT words ($r=.598, p=0.02$). MRT performance was also correlated with performance on Anomalous sentences ($r=.515, p=0.008$) and Environmental stimuli ($r=.554, p=0.004$). 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.

between materials drawn from the pre-test and novel list ($p=0.477$), indicating that training generalized to new words of the same class, and that performance was not contingent on having heard the word before.

Anomalous Sentence Training. Figure 4 shows subject performance on the Anomalous sentence training condition. Performance was good at pre-test (33.6% correct) but increased significantly following training (61.7% correct, $t(1,24)=11.713$, $p<0.001$). Examination of the post-test materials (data not shown) revealed a significant main effect of source ($F(2, 74)=14.115$, $p<0.001$), and post hoc Tukey tests confirmed that subjects performed significantly better on the materials from the training list (78.2% correct) than on materials from either the pre-test list (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 ($p=0.998$).

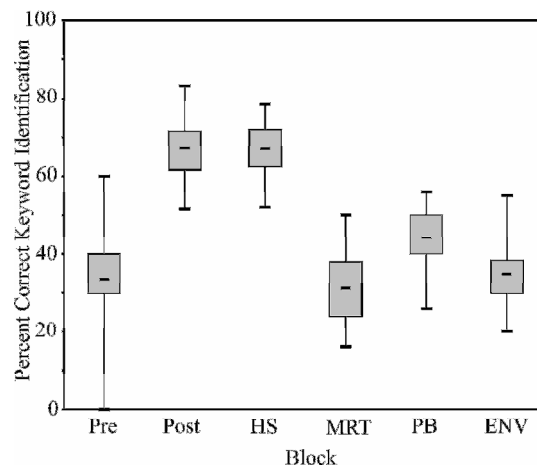


FIGURE 4. Box plots displaying the perceptual accuracy scores as a function of experimental block for the 25 subjects trained to identify the Anomalous sentences.

As observed previously, subjects performed best on the Harvard/IEEE sentences (67.04% correct), followed by the anomalous sentences (61.7% correct), PB words (44.1% correct), environmental stimuli (34.9% correct) and MRT words (31.2% correct). Performance on the Harvard/IEEE sentences was significantly higher than on the Anomalous sentences ($t(1,24)=3.406$, $p=0.002$), and subtraction of the scores on these blocks revealed only a small 5% gain from context, suggesting that the large gains due to context observed in the MRT and PB training were ameliorated with explicit training on the anomalous sentences. Subjects also performed significantly better on PB as compared to MRT words ($t(1,24)=6.140$, $p<0.001$), as observed previously. Performance on the Anomalous sentences was correlated only with performance on Harvard/IEEE sentences ($r=.63$, $p=0.001$). The only other significant correlation observed was between PB words and Environmental stimuli ($r=.446$, $p=0.025$). All other correlations were not significant.

Harvard/IEEE Sentence Training. Performance on the Harvard/IEEE sentence post-test significantly increased from pre (40% correct) to post-test (63.9% correct, $t(1,24)=7.041$, $p<0.001$). Subject performance varied across the post-test materials, and an ANOVA analysis revealed a significant main effect of source ($F(2, 74)=114.043$, $p<0.001$). Subjects performed significantly better on materials from the training list (97% correct) than on those from the pre-test (71.6% correct) and novel (56.2%

correct) lists (all $p < 0.001$). Subjects also performed significantly better on the 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.

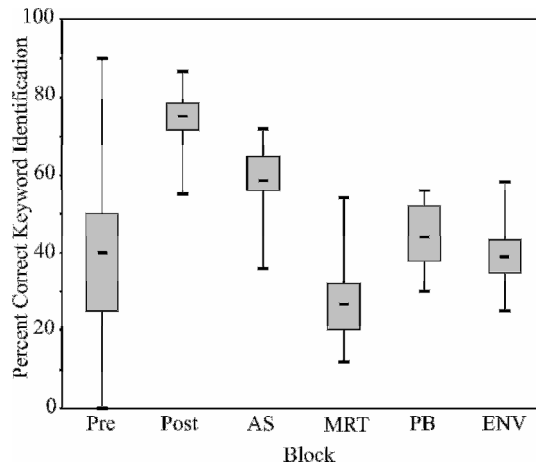


FIGURE 5. Box plots displaying the perceptual accuracy scores as a function of experimental block for the 25 subjects trained to identify the Harvard/IEEE sentences.

Figure 5 shows that subjects performed best on the Harvard/IEEE sentences, followed by the Anomalous sentences (58.6% correct), PB words (44.1% correct), Environmental stimuli (39% correct) and MRT words (26.7% correct). A paired t-test revealed that subjects performed significantly better on the Harvard/IEEE sentences than on the Anomalous sentences ($t(1,24)=3.328, p=0.003$). The gain from context was only approximately 5%. Subjects also performed significantly better on the PB words as compared to the MRT words ($t(1,24)=10.332, p < 0.001$). Performance on the Harvard/IEEE sentences was significantly correlated with performance on Anomalous sentences ($r=.551, p=0.004$), followed by PB words ($r=.512, p=0.009$) and MRT words ($r=.398, p=0.49$). Anomalous sentences were most strongly correlated with performance on PB words ($r=.733, p < 0.001$) and MRT words ($r=.623, p=0.001$). Performance on PB words was significantly correlated with performance on MRT words ($r=.587, p=0.002$) and Environmental stimuli ($r=.568, p=0.003$).

Environmental Stimulus Training. Performance on the Environmental stimuli also showed a significant benefit from explicit training (Fig. 6). Subjects showed significant improvement between pre (38.2% correct) and post-test (46.4% correct, $t(1,24)=2.804, p=0.01$). An analysis of the post-test materials (data not shown) revealed a significant main effect of source ($F(2, 74)=8.717, p < 0.001$). Subjects performed best on stimuli from the novel list (53.2% correct), followed by materials from the training list (50% correct) and pre-test (39.6% correct). Subjects performed significantly better on materials from both novel and training lists than on materials from the pre-test list ($p=0.009$ and $p < 0.001$ respectively) but did not differ from one another ($p=0.617$).

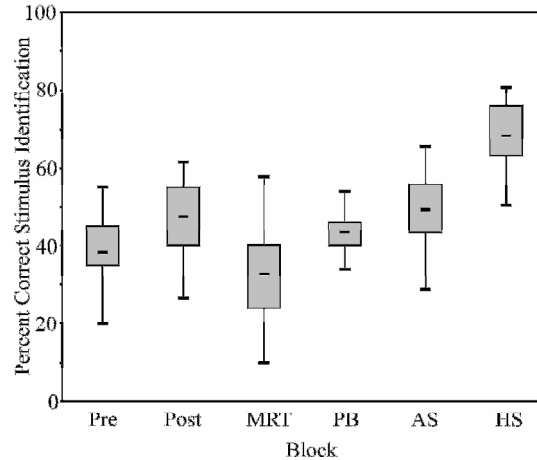


FIGURE 6. Box plots displaying the perceptual accuracy scores as a function of experimental block for the 25 subjects trained to identify the Environmental stimuli.

Overall, subjects performed best on Harvard/IEEE sentences (68.5% correct), followed by the Anomalous sentences (49.31% correct), Environmental stimuli, PB (43.4% correct) and MRT words (32.8% correct). Subjects received an approximated gain of 19% from context ($t(1,24)=13.772$, $p<0.001$). Subjects also performed significantly better on PB words as compared to MRT words ($t(1,24)=3.830$, $p=0.001$). Performance on the Environmental stimuli was not significantly correlated with any other material, but as observed earlier, Harvard/IEEE sentences were significantly correlated with Anomalous sentences ($r=.69$, $p<0.001$).

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 and to the scores at pre-test. Comparison with the post-test scores (which did not contain the materials repeated from pre-test) 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.

MRT Words. Figure 7 displays the across group performance on the MRT words. A one-way ANOVA using Training Materials as the between subjects factor main effect of training materials ($F(5, 149)=37.495$, $p<0.001$). Post hoc Tukey tests revealed that subjects performed significantly better than the pre-test 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, MRT and PB training produced greater benefits than any other material (37.5% correct, and 42.8% correct respectively). That performance did not differ between the MRT and PB trained groups ($p=0.477$) suggests that training on words, regardless of their origin, produces equivalent benefit when recognizing other single words. Training on Anomalous sentences, Harvard/IEEE sentences and Environmental stimuli also produced significant gains over baseline, but were the poorest of all conditions (31.2% correct, 26.7% and 32.8% correct respectively). Moreover, performance did not differ between these three groups (all $p>0.319$). Interestingly, subjects trained on the Anomalous sentences and Environmental 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 stimuli ($p=0.027$).

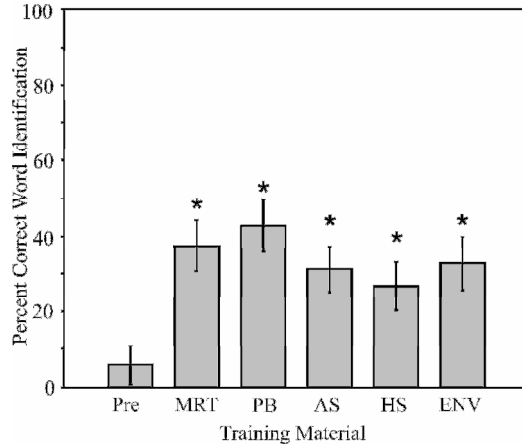


FIGURE 7. Bar graph displaying perceptual accuracy scores at identifying MRT stimuli as a function of training. Training condition is indicated along the x-axis. Pre and post-test scores are for the subjects who were explicitly trained on the MRT stimuli. The remaining bars indicate subjects’ performance on the MRT generalization block of their respective training sessions. Post-test scores contain only the responses to stimuli on which subjects did not receive explicit training (see text). Asterisks indicate when performance was significantly greater than baseline ($p < 0.05$). stimuli performed as well as subjects trained on the MRT stimuli ($p = 0.281$ and $p = 0.610$ respectively).

PB Words. Training produced a significant impact on subjects performance on the PB materials (Figure 8), and a one-way ANOVA using Training Materials as the between subjects factor indicated a significant main effect materials ($F(5, 149) = 24.86, p < 0.001$). Compared to baseline, post-test performance is significantly higher as a result of training on PB materials (23.4% as compared to 46.2%, $p < 0.001$). Overall, it did not matter what type of training subjects received, as performance was significantly higher than pre-test for all training conditions (MRT training 43.4% correct $p < 0.001$, AS training 44.1% correct $p < 0.001$, HS training 44.1% 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 pre-test, as there were no significant differences between performance across the five training conditions (all $p > 0.867$). This indicates that when identifying words that are highly discriminable, training with any type of material will provide an equivalent benefit.

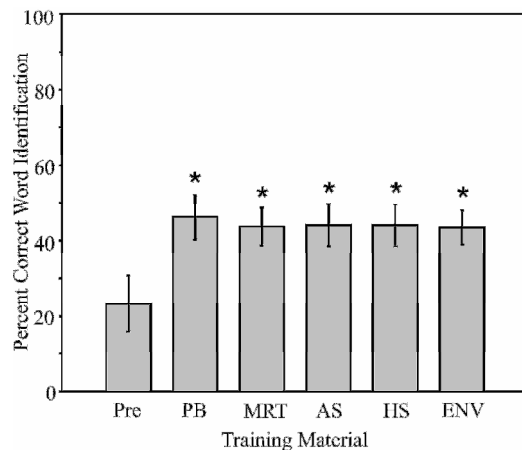


FIGURE 8. Bar graph displaying perceptual accuracy scores at identifying PB stimuli as a function of training. Training condition is indicated along the x-axis. Pre and post-test scores are for the subjects who were explicitly trained on the PB stimuli. The remaining bars indicate subjects’ performance on the PB generalization block of their respective training sessions. Post-test scores contain only the responses to stimuli on which subjects did not receive explicit training (see text).

Anomalous Sentences. The performance on the anomalous sentences across training conditions is shown in Figure 9. A one-way ANOVA revealed a significant main effect of training ($F(5, 149) = 22.986, p < 0.001$). Comparison with the pre-test revealed that all types of training produced significant increases in performance relative to the baseline (33.6% correct, all $p < 0.001$). No differences in performance were observed between subjects who received explicit training on the Anomalous sentences (61.7% correct) and to those who were trained on the meaningful Harvard/IEEE sentences (58.6% correct, $p = 0.902$). In contrast, subjects who received training on the PB, MRT and Environmental stimuli showed significantly less gain in performance as compared to subjects trained on the Anomalous (all $p < 0.001$) or Harvard/IEEE (all $p < 0.004$) sentences. Training on MRT (47.7% correct), PB (46.5% correct) and Environmental stimuli (47.7% correct) provided equivalent benefit when recognizing the Anomalous sentences, however (all $p > 0.0998$).

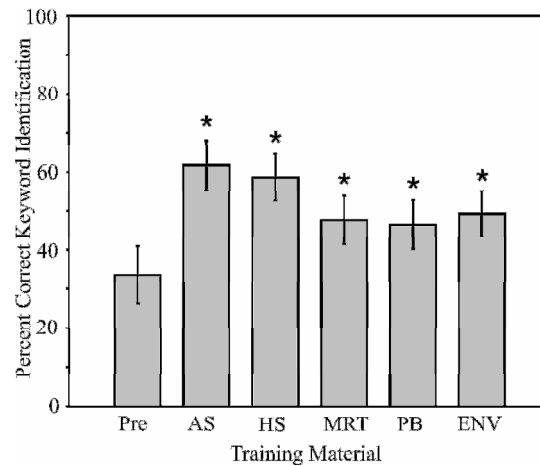


FIGURE 9. Bar graph displaying perceptual accuracy scores at identifying Anomalous sentences as a function of training. Training condition is indicated along the x-axis. Pre and post-test scores are for the subjects who were explicitly trained on the Anomalous sentences. The remaining bars indicate subjects' performance on the AS generalization block of their respective training sessions. Post-test scores contain only the responses to stimuli on which subjects did not receive explicit training (see text).

Harvard/IEEE Sentences. The comparison of performance on the Harvard/IEEE sentences across training conditions is shown in Fig. 10. A one-way ANOVA revealed a significant main effect of training condition on performance ($F(5, 149) = 22.444, p < 0.001$). The comparison of each of the training conditions to the Harvard/IEEE sentence pre-test revealed that 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 is carried entirely by the gains in performance relative to the pre-test, as there were no significant differences between performance across the five training conditions (MRT 67.0% correct, PB 66.2% correct, HS 63.9% correct, AS 67.0% correct, ENV 68.5% correct, all $p > 0.719$).

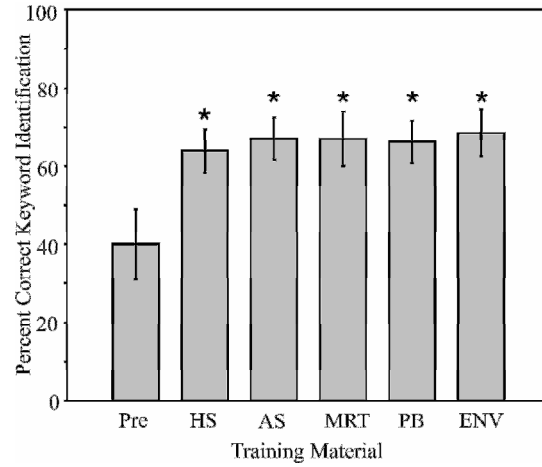


FIGURE 10. Bar graph displaying perceptual accuracy scores at identifying Harvard/IEEE sentences as a function of training. Training condition is indicated along the x-axis. Pre and post-test scores are for the subjects who were explicitly trained on the Harvard/IEEE sentences. The remaining bars indicate subjects' performance on the HS generalization block of their respective training sessions. Post-test scores contain only the responses to stimuli on which subjects did not receive explicit training (see text).

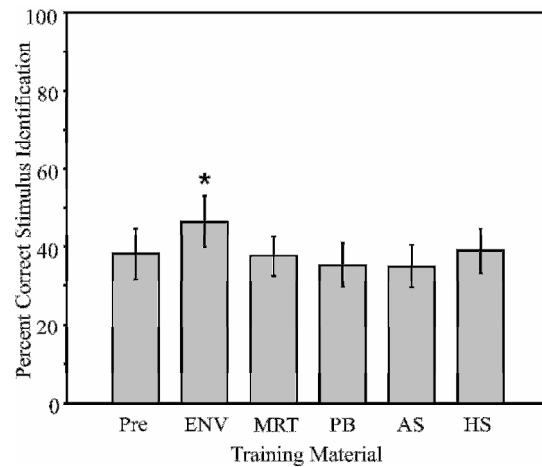


FIGURE 11. Bar graph displaying perceptual accuracy scores at identifying the Environmental stimuli as a function of training. Training condition is indicated along the x-axis. Pre and post-test scores are for the subjects who were explicitly trained on the Environmental stimuli. The remaining bars indicate subjects' performance on the ENV generalization block of their respective training sessions. Post-test scores contain only the responses to stimuli on which subjects did not receive explicit training (see text).

Environmental Stimuli. The effect of training on the recognition of the Environmental stimuli is shown in Figure 11. A one-way ANOVA revealed a significant main effect of training group on performance ($F(5, 149) = 5.847, p < 0.001$). 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 stimuli (46.4% 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 34.9% correct $p=0.999$; AS 34.9% correct $p=0.764$). Moreover, performance was significantly higher for those subjects explicitly trained on the Environmental stimuli as compared to all other groups (all $p<0.03$). Training on MRT, PB, AS and HS materials provided equivalent levels of generalization to the Environmental stimuli (all $p>0.557$). Since these values did not differ from the baseline, however, it suggests that training on the speech materials is equally ineffective when transferring to environmental stimuli. In effect, when asked to identify environmental stimuli, training with speech materials is as effective as not receiving any training at all.

Discussion

Overall, the specific type of materials used during the training portion of the experiment had a significant impact on performance. Across all training conditions, subjects showed significant pre to post-test improvement, demonstrating that for each set of training materials, subjects were able to utilize the feedback to improve their identification accuracy. Generalization effects were not uniform across materials. Subjects showed encoding specificity, performing best on the materials on which they were explicitly trained. Subjects who were trained on words (PB or MRT) performed significantly better when identifying MRT stimuli than the other groups, and subjects who were trained on sentences (anomalous or meaningful) performed significantly better when identifying Anomalous sentences than the other groups. This suggests that when the task demands were high, 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 of appropriate processing. The opposite effect was observed for the “easier” 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 high frequency words and meaningful sentences, all forms of training are equivalent.

One intriguing finding from this study was the asymmetry in training that was observed for the environmental stimuli. Subjects trained on environmental stimuli performed significantly better than baseline on all speech materials, suggesting that training on complex non-speech 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. Thus, it appears that 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. Increased attentional sensitivity to the spectral and temporal characteristics of the environmental stimuli may have enhanced subjects’ abilities at utilizing similar spectral information that is important to speech.

The present findings are similar to those of Gygi and colleagues, who found that the most important information for recognition of environmental stimuli occupies an identical frequency range as that for speech (Gygi et al., 2004). If the important information for environmental stimuli overlaps with that of speech, then training subjects to better utilize the spectro-temporal information in this frequency region more efficiently should foster generalization to speech, as we report here. Training on speech alone may not be sufficient to foster generalization to environmental stimuli, since the spectro-temporal information to which subjects are utilizing may be more broadly distributed for these stimuli. Additionally, some environmental stimuli may be inherently more identifiable than others based on their spectro-temporal profiles (Shafiro, 2004; Burkholder, 2005; Burkholder et al., submitted 1; 2). The interaction between the number of spectral bands needed for successful recognition that was found by Shafiro (2005) was somewhat divergent from that typically observed for speech. Some environmental stimuli were most recognizable with fewer bands, and recognition actually decreased with the addition of bands (Shafiro, 2005). This suggests that some environmental stimuli may not be as readily identifiable

when processed by a vocoder. Moreover, given that the amount of acoustic information differs across acoustic environments and task demands, the spectral resolution of the current generation of CIs may be insufficient to provide significant benefit under all listening situations (Shannon, Fu & Galvin, 2004; Shannon, 2005). This possibility warrants further investigation.

The finding that training on speech does not generalize to environmental stimuli conflicts with the earlier findings of Burkholder and colleagues (Burkholder, 2005; Burkholder et al., submitted 1; 2), who reported that training on speech did generalize to environmental stimuli. However, Burkholder did not use a pre-post test design, so the baseline performance levels for environmental stimuli were not known. In the present study, although training on speech materials produced performance levels for environmental stimuli that were greater than zero, they did not exceed the baseline values. This suggests that subjects in the Burkholder et al study (submitted 1, 2) may not have performed any differently after training than subjects who were totally naïve to the stimulus processing conditions.

One methodological difference between the present study and earlier studies using environmental stimuli is the use of open set testing procedures in all conditions. The majority of the earlier studies used closed-set forced-choice testing procedures. Gygi and colleagues reported closed-set identification scores of up to 66% correct using 6-channel noise vocoded stimuli (Gygi et al., 2004). Shafiro found that although closed-set performance reaches asymptote with 16 channels (66%), large stimulus specific effects were observed (Shafiro, 2004). Moreover, Reed and Delhorne (2005) found that CI users show higher levels of closed-set performance still (79% correct). 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, Pisoni & Tierney, 2006 for a more complete account).

A methodological question is also raised here. Although many previous studies have not demonstrated 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 & Oliver, 2005). Gender and talker identification were significantly better for stimuli processed using a sinewave vocoder than when processed using a noise vocoder (Gonzales & 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 & Oliver, 2005). It may be the case that a sinewave vocoder may produce better, more robust results for studies using music and environmental stimuli 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 differs from performance of CI users remains an open question.

The asymmetry in training that was observed in the present study suggests that the ability to utilize the residual spectro-temporal information in the vocoded signals may enhance the ability to perceive unfamiliar speech signals under these difficult listening conditions. Surprenant and Watson (2001) reported a significant correlation between subjects' ability to discriminate non-speech stimuli based on spectro-temporal cues and their identification of speech in noise. The authors suggested that common higher order acoustic processes may contribute to both speech and non-speech processing capabilities. This could account for the substantial differences in performance of subjects who receive

hearing aids, and CIs alike: auditory sensitivity at a peripheral level may not be the sole cause of variability; rather the inability to utilize and manipulate such information at higher levels may supersede the benefits of an acoustic prosthesis (Surprenant & Watson, 2001). This relationship may not be completely bidirectional, however, given our findings that training on environmental stimuli generalizes to speech, but training on speech does not generalize to environmental stimuli.

Moreover, recent neuroimaging studies investigating the encoding of environmental stimuli have suggested that similar cortical regions may be involved during the processing of environmental stimuli and speech sounds (Lewis, Wightman, Brefczynski, Phinney, Binder & DeYoe, 2004). These cortical regions include the 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, pSTS), 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 stimuli, 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 stimuli and speech. Perhaps training with environmental stimuli 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 which would not necessarily facilitate generalization to environmental stimuli. 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, Wise, Dresner & Scott, 2007). This may facilitate generalization in one case (environmental stimuli to speech), but not the other (speech to environmental stimuli).

Our findings also replicate and extend the recent studies conducted by Davis et al (2005) and Burkholder et al (submitted 1, 2). Training using orthographic feedback paired with a repetition of the processed version of the sentence produced keyword correct identification scores (71% correct) that were nearly identical to those observed 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 et al. (submitted 1, 2). Thus, access to syntactic structure without relying on sentence context enhances general sentence recognition. Our extension to include single PB words and CVCs also provides support for this conclusion: training on all materials produced excellent generalization to the meaningful Harvard/IEEE sentences. The results observed for training on environmental stimuli suggest that learning to recognize the acoustic form of a stimulus enhances selective attention to spectro-temporal information, and bottom up perceptual encoding processes.

The present study also replicates 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 CI 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 behavioral 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 CI 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 CI users that cannot be

attributed to individual differences in etiology and duration of deafness, the question remains as to how differences in post-implantation 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 the CI user success.

In summary, we demonstrated that the type of stimulus materials used during perceptual learning affects generalization to new materials. Although all forms of training provided some benefit, generalization of training was not uniform. 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 difficult, such as was the case when identifying low discriminable CVCs or sentences without the benefit of 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 signals revealed a unique asymmetry: training on environmental signals generalized to the recognition of speech, but training on speech did not generalize to environmental signals. 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 signals.

References

- Bradlow, A.R., Toretta, G.M. and Pisoni, D.B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, 20, 255-272.
- Burkholder, R.A. (2005). Perceptual learning of speech processed through an acoustic simulation of a cochlear implant. *Research on Spoken Language Processing Technical Report No. 13*, Bloomington, IN: Speech Research Laboratory, Indiana University.
- Burkholder, R.A., Pisoni, D.B. and Svirsky, M.A. (submitted 1). Transfer of auditory perceptual learning with spectrally reduced speech to speech and nonspeech tasks: Implications for cochlear implants. *Ear and Hearing*.
- Burkholder, R.A., Pisoni, D.B. and Svirsky, M.A. (submitted 2). 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*.
- Chiu, C.-Y.P., and Schacter, D.L. (1995). Auditory priming for nonverbal information: Implicit and explicit memory for environmental sounds. *Consciousness and Cognition*, 4, 440-458.
- Chiu, C.-Y.P., (2000). Specificity of auditory implicit and explicit memory: Is perceptual priming for environmental sounds exemplar specific? *Memory and Cognition*, 28(7), 1126-1139.
- Clark, G.M. (2002). Learning to understand speech with the cochlear implant, In Fahle, M, and Poggio, T. Eds. *Perceptual Learning*, pp. 147-160. Boston: MIT press.
- Clopper, C.G., Pisoni, D.B. and Tierney, A.T. (2006). Effects of open-set and closed-set task demands on spoken word recognition. *Journal of the American Academy of Audiology*, 17(5), 331-349.
- Davis, M.H., Johnsruide, I.S., Hervais-Adelman, A., Taylor, K. and McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise vocoded sentences. *Journal of Experimental Psychology*, 134(2), 222-241.
- Dorman, M.F., Loizou, P.C. and Rainey, D. (1997). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *Journal of the Acoustical Society of America*, 102(1), 2993-2996.
- Dorman, M. and Loizou, P. (1998). 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 and Hearing*, 19, 162-166.

- Egan, J.P. (1948). Articulation testing methods. *Laryngoscope*, 58, 955-991.
- Fu, Q.-J., Galvin, J., Wang, X. and Nogaki, G. (2006). Moderate auditory training can improve speech performance of adult cochlear implant patients. *Acoustic Research Letters Online*, 6(3), 106-111.
- Gonzales, J. and Oliver, J.C. (2005). Gender and speaker identification as a function of the number of channels in spectrally reduced speech. *Journal of the Acoustical Society of America*, 118, 461-470.
- Gygi, B., Kidd, R.R. and Watson, C.S. (2004). Spectral-temporal factors in the identification of environmental sounds. *Journal of the Acoustical Society of America*, 115(3), 1252-1265.
- Herman, R. and Pisoni, D.B. (2000). Perception of elliptical speech by an adult hearing-impaired listener with a cochlear implant: some preliminary findings on coarse-coding in speech perception. *Research on Spoken Language Processing Progress Report No. 24* (pp. 87-112) Bloomington, IN: Speech Research Laboratory, Indiana University.
- House, A.S., Williams, C.E., Hecker, M.H.L. and Kryter, K.D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America*, 37, 158-66.
- IEEE. (1969). IEEE recommended practice for speech quality measurements. (IEEE Report No. 297).
- Karl, J.R. and Pisoni, D.B. (1994). Effects of stimulus variability on recall of spoken sentences: A first report. *Research on Spoken Language Processing Progress Report No. 19* (pp. 145-193) Bloomington, IN: Speech Research Laboratory, Indiana University.
- Lachs, L., McMichael, K. and Pisoni, D.B. (2003). Speech perception and implicit memory: Evidence for detailed episodic encoding of phonetic events. In J. Bowers and C. Marsolek (Eds.), *Rethinking implicit memory*. (pp. 215-235). Oxford: Oxford University Press.
- Lewis, J.W., Wightman, F.L., Brefczynski, J.A., Phinney, R.E., Binder, J.R. and DeYoe, E.A. (2004). Human brain regions involved in recognizing environmental stimuli. *Cerebral Cortex*, 14(9), 1008-1021.
- Marcell, M.M., Borella, D., Greene, M., Kerr, E., and Rogers, S. (2000). Confrontation naming of environmental sounds. *Journal of Clinical and Experimental Neuropsychology*, 22(6), 830-864.
- Miller, G.A. and Nicely, P. (1955). An Analysis of Perceptual Confusions among some English Consonants, *Journal of the Acoustical Society of America*, 27(2), 338-352.
- National Institutes of Health. (1995). Cochlear implants in adults and children. *NIH Consensus statement*, 13(2), 1-29.
- Obleser, J., Wise, R.J.S., Dresner, M.A., and Scott, S.K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience*, 27(9), 2283-2289.
- Reed, C.M and Delhorne, L.A. (2005). Reception of environmental sounds through cochlear implants. *Ear and Hearing*, 26(1), 48-61.
- Shafiro, V. (2004). Perceiving the sources of environmental sounds with a varying number of spectral channels, Unpublished doctoral dissertation. CUNY.
- Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J. and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 303-304.
- Shannon, R.V., Fu, Q.-J. and Galvin, J. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Otolaryngologica Supplementum*, 552, 1-5.
- Shannon, R.V. (2005). Speech and music have different requirements for spectral resolution. *International Review of Neurobiology*, 70, 121-134.
- Stevens, K.N. (1980). Acoustic correlates of some phonetic categories. *Journal of the Acoustical Society of America*, 68(3), 836-842.

- Surprenant, A.M. and Watson, C.S. (2001). Individual differences in the processing of speech and nonspeech sounds by normal hearing listeners. *Journal of the Acoustical Society of America*, *110*(4), 2085-2095.
- Tye-Murray, N., Tyler, R., Woodward, G. and Gantz, B. (1992). Performance over time with a Nucleus and Ineraid cochlear implant. *Ear and Hearing*, *13*(3), 200-209.