

**RESEARCH ON SPOKEN LANGUAGE PROCESSING**

Progress Report No. 28 (2007)

*Indiana University*

**Multiple Routes to Perceptual Learning<sup>1</sup>**

**Jeremy L. Loebach, Tessa Bent, and Althea Bauernschmidt**

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

---

<sup>1</sup> This research supported by NIH NIDCD R01 Research Grant DC00111, and NIH NIDCD T32 Training Grant DC00012 to Indiana University. The authors wish to thank Larry Phillips for his assistance in data collection. We would also like to thank Luis Hernandez for providing technical assistance and advice in the design and implementation of the experimental procedures.

## Multiple Routes to Perceptual Learning

**Abstract.** A listener's ability to utilize indexical information in the speech signal can enhance their performance on a variety of speech perception tasks. It is unclear, however, whether such information plays a similar role for spectrally reduced speech signals, such as those experienced by individuals with cochlear implants. The present study compared the effects of training on linguistic versus indexical tasks when adapting to cochlear implant simulations. Listening to sentences processed with an 8-channel sinewave vocoder, three groups of subjects were trained on a transcription task (Transcription), a talker identification task (Talker ID) or a gender identification task (Gender ID). Pre- to post-test comparisons demonstrated that training produced significant improvement for all groups. Moreover, subjects from the Talker ID and Transcription training groups performed similarly at post-test and generalization, and significantly better than the subjects from the Gender ID training group. These data suggest that training on an indexical task that requires high levels of attention can provide equivalent benefit to training on a linguistic task. When listeners selectively focus their attention on the extra-linguistic information in the speech signal, they still extract linguistic information, the degree to which they do so, however, appears to be task dependent.

### Introduction

The acoustic speech stream contains two different sources of information: linguistic information, which carries the meaning of the utterances, and indexical information, which specifies the characteristics of the speaker's voice (e.g. gender, age, dialect) (Ladefoged & Broadbent, 1957). How these two types of information interact during speech processing is largely unknown. Does the listener encode linguistic and indexical information in independent streams via different perceptual mechanisms, or are they encoded and processed together? The present study addressed this question by investigating how selectively focusing the listener's attention on linguistic or indexical information during training affects adaptation to spectrally degraded speech. Using sentences that had been processed by a cochlear implant (CI) simulator, we investigated how different types of training affected both perceptual learning and generalization to new sentences, talkers, and more severely spectrally degraded conditions. We found that the amount of attention required during the training task modulated the relative gain and strength of perceptual learning. Training on Talker ID, an indexical task that required a higher degree of attentional control and focus on the acoustic information in the signal, elicited more robust generalization than training on Gender ID.

### Indexical Information Enhances Linguistic Processing

Indexical characteristics of talkers are important for successful interpersonal communication. A talker's particular realizations of acoustic-phonetic parameters will ultimately determine their intelligibility (Bond & Moore, 1994; Bradlow, Toretta & Pisoni, 1996; Cox, Alexander & Gilmore, 1987; Hood & Poole, 1980). Adaptation to talker idiolect is a natural part of speech perception, and adult listeners are constantly adjusting their internal categories to accommodate new talkers. Such perceptual learning, which can be defined as long-term changes in the perceptual system based on sensory experience that will influence future behaviors and responses (Goldstone, 1998; Fahle & Poggio, 2002), may play a central role in adaptation to novel talkers. When a listener is explicitly trained to classify an ambiguous sound in a word in which it does not belong (such as the word "vacation" produced with a /z/ versus a /s/), category boundaries for words containing the sound will be adjusted to accommodate the

new pronunciation (Eisner & McQueen, 2005). This result only holds if the talker used during training is included in the test set, however (Eisner & McQueen, 2005). In this case, the phonemic distinction is relatively isolated, and listeners do not generalize to new talkers.

In addition, familiarity with a talker's voice can enhance speech perception under difficult listening conditions (Nygaard, Sommers & Pisoni, 1994). Listeners trained to identify talkers by name demonstrated better word identification accuracy than listeners who were unfamiliar with the test talkers (Nygaard et al., 1994; Nygaard & Pisoni, 1998). Two distinct types of subjects were observed: "good" learners, who exceeded 70% correct talker identification and "poor" learners, who did not (Nygaard & Pisoni, 1998). "Poor" learners performed significantly worse on word and sentence identification after training than did the "good" learners, suggesting that it is not the mere exposure to the talkers that is enhancing word identification accuracy, but rather the ability to store and utilize the acoustic information that characterize the talker's voice. When taken together, these data demonstrate the presence of significant interactions between the linguistic and indexical channels of information in speech, and suggest that the two may indeed be coded in the same stream.

Listeners can adapt not only to specific talkers but given the appropriate exposure also show talker-independent adaptation to talkers from a variety of special populations whose speech deviates from normal, native talker norms. For example, when first confronted with a non-native speaker, many listeners may have difficulty understanding them, but with exposure, they quickly learn to adapt to their speaking patterns (Bradlow & Bent, in press; Clarke & Garrett, 2004; Weil, 2001). Similarly, a beneficial effect of experience on speech intelligibility has been shown for listeners with extensive experience listening to speech produced by talkers with hearing impairments (McGarr, 1983), computer manipulated speech (Schwab, Nusbaum & Pisoni, 1985; Greenspan, Nusbaum & Pisoni, 1988; Dupoux & Green, 1997; Pallier, Sebastian-Gallés, Dupoux, Christophe & Mehler, 1998), and noise-vocoded speech (Davis, Johnsrude, Hervais-Adelman, Taylor & McGettigan, 2005). Critically, this benefit extends to new talkers, or to new speech signals created using the same types of signal degradation.

Furthermore, adaptation to a talker's idiolect may not be completely talker specific, however, if the training contrasts are lexically contrastive in the language and have a greater degree of potential generalizability (Kraljic & Samuel, 2006). When exposed to words containing an ambiguous sound between /d/ and /t/ in which the voicing distinction is blurred (e.g., "crocatile" or "cafederia"), subjects show robust generalization to novel utterances containing the ambiguous phoneme produced by novel talkers. Moreover, perceptual learning generalizes to a novel consonant set including an ambiguous /b/ - /p/ in which the voice onset time boundary is similarly blurred. These data suggest that when the phonemic distinction is important to more phonemes than are used in the training set (as is the case for the voicing distinction), generalization will be robust and occur independent of talker.

Compared to the literature on the perceptual learning of naturally produced speech, the explicit perceptual learning and generalization of spectrally reduced speech has received little attention. Previous research using sinewave speech has demonstrated that subjects trained to identify talkers from sentences containing three sinewave analogs of the formant frequencies show robust generalization when asked to identify these same talkers from naturally produced versions of the sentences (Remez, Fellowes & Rubin, 1997; Sheffert, Pisoni, Fellowes & Remez, 2002). The effect is not bidirectional, however, since training with naturally produced speech does not generalize to sinewave speech (Sheffert et al., 2002). These data suggest that talker identification of sinewave speech may be utilizing different acoustic information than is normally used for talker identification of naturally produced speech. Since sinewave speech is derived from natural speech, some acoustic cues will be shared in both types of stimuli, promoting generalization from sinewave to naturally produced speech. When trained on naturally produced speech, however, the

listener may rely on other acoustic cues that are not preserved in the sinewave analogs, and as such, generalization to the sinewave utterances does not occur (Sheffert et al., 2002). Thus, it appears that the listener is opportunistic, relying on whatever acoustic cues are available in the signal in order to identify the talker. Although the findings with sinewave speech demonstrate that talker identification training with spectrally reduced speech generalizes to talker ID tasks for naturally produced speech, these studies have not assessed whether training on talker identification generalizes to word or sentence recognition under conditions of severe spectral degradation as has been shown for naturally produced speech (Nygaard & Pisoni, 1998).

The type of training a listener receives during adaptation to spectrally degraded speech affects the extent of perceptual learning and transfer to new materials. Feedback promotes more rapid adaptation to CI simulated speech than no feedback (Davis, et al., 2005). Moreover, the type of feedback that is given can also modulate the speed of perceptual learning. The most effective feedback includes the processed audio stimuli paired with the orthographic representation (Burkholder, 2005). Additionally, training with complex non-speech environmental stimuli promotes transfer and generalization to speech materials (Loebach & Pisoni, under review). Whether training on indexical tasks will generalize to speech perception under CI simulations, however, is unknown.

### **Indexical Information in Cochlear Implants**

Although cochlear implants have been successful in providing the profoundly hearing impaired with access to the acoustic signal, a large amount of variability remains among cochlear implant users. While the age at onset of deafness, duration of auditory deprivation and etiology of deafness all influence outcomes after implantation, these factors do not account for all intra-subject variability (NIH, 1995). Moreover, research with CI users has focused almost exclusively on speech perception, leaving the perception of other types of acoustic signals (e.g., meaningful environmental sounds) unexplored. Although ideally individuals will achieve high levels of speech perception in quiet and noise, not all CI users will receive such a benefit. At a minimum, the individual is expected to gain some awareness of sound, including environmental stimuli (Clark, 2002).

For linguistic tasks, acoustic simulations of cochlear implants have provided a useful tool for determining what acoustic information is necessary for speech perception. Early work demonstrated that sufficient linguistic information is conveyed via acoustic simulations of a cochlear implant processor and electrode array to allow the identification of single consonants, vowels and sentences (Shannon, Zeng, Kamath, Wygonski & Ekelid, 1995). Designed to simulate different numbers of active electrodes in the intracochlear array, these simulations have demonstrated that successful speech perception is largely dependent on number of acoustic channels. Under quiet listening conditions, normal hearing subjects reach asymptote for sentences containing eight channels (Dorman, Loizou & Rainey, 1997), although more channels are needed when listening in noise (Dorman, Loizou, Fitzke & Tu, 1998). Furthermore, normal hearing subjects listening to 6 channel simulations perform similarly to cochlear implant users (Dorman & Loizou, 1998). Although limited spectral information is sufficient for high levels of consonant, vowel and sentence perception, other tasks may require substantially more spectral information. Acoustic stimuli that contain complex acoustic spectra, such as music, may require well over thirty channels to be perceived accurately (Shannon, Fu & Galvin, 2004; Shannon, 2005).

Compared to perception of linguistic information in the speech signal, considerably less is known about the perception of indexical information both in CI users, and in normal hearing subjects listening to CI simulations. Cleary and Pisoni (2002) demonstrated that prelingually deafened children with cochlear implants have more difficulty discriminating talkers based on their voices than do normal hearing

children. Moreover, considerable variability existed across subjects: over half of the children who had cochlear implants could not discriminate talkers at a level greater than chance, while those who could discriminate talkers performed comparably to the normal hearing children (Cleary, Pisoni & Kirk, 2005). When considered as a group, all children with cochlear implants required larger pitch deviations between talkers in order to distinguish them, and showed more pronounced difficulty in talker discrimination when the sentences varied across talkers than did normal hearing children (Cleary et al., 2005).

While talker discrimination may rely on acoustic details that are not well conveyed by a cochlear implant processor, gender discrimination may utilize primarily temporal cues. Normal hearing subjects listening to CI simulations require more spectral channels to accurately discriminate the gender of talkers than to identify vowels from a closed set response set (Fu, Chinchilla & Galvin, 2004). As the number of channels increase from four to thirty-two, percent correct gender identification increased approximately linearly. Moreover, a tradeoff between spectral and temporal information was observed for gender discrimination: fewer spectral channels are required when more precise temporal information is preserved. CI users' performance was roughly comparable to normal hearing subjects listening to four or eight band simulations. Moreover, accuracy depends on the individual voices of the talkers who are used in the study. If the differences between male and female talkers are large, normal hearing subjects and CI users utilize temporal information to classify the speakers' gender based on their fundamental frequency (Fu, Chinchilla, Nogaki & Galvin, 2005). Thus, it appears that CI users may be relying primarily on temporal pitch information to distinguish talkers, a strategy that becomes ineffective when the difference between male and female fundamental frequencies decreases (Fu et al., 2005).

The performance on gender identification tasks is also dependant on the method of synthesis. While speech perception accuracy does not differ for noise and sinewave vocoders (Dorman et al., 1997), gender discrimination is more accurate with sinewave than noise vocoders (Gonzalez & Oliver, 2005). Compared to noise vocoders, subjects listening to sinewave vocoders require fewer channels to reach asymptote on the gender identification task.

Gender identification and talker discrimination, however, require different types of processing compared to talker identification. The acoustic cues that allow the listener to discriminate male from female talkers or to decide if two sentences are produced by the same or different talkers may be much coarser than those required to identify a speaker from their voice alone. Vongphoe and Zeng (2005) trained normal hearing subjects and CI users to identify ten talkers and compared talker and vowel identification accuracy. Normal hearing subjects listening to sinewave vocoded vowels achieved high levels of talker identification accuracy, particularly with stimuli containing more spectral channels (e.g., 32 channels). Cochlear implant users performed significantly worse than the normal hearing listeners. For vowel recognition, however, performance by CI users approximated the normal hearing subjects listening to 8-channel vocoders. The differences in performance of the CI users on the vowel and talker identification tasks led the authors to conclude that the subjects may be utilizing different processing strategies during linguistic and indexical tasks (Vongphoe & Zeng, 2005).

One possible confound in the study, however, comes from the overlap in the fundamental frequencies of the talkers voices. When considered on a talker-by-talker basis the predominant source of errors in talker identification was not between adult male and adult female talkers, but from confusions between the voices of adult females, girls and boys (Vongphoe & Zeng, 2005). Given that the dominant confusions were between talkers with higher pitched voices, the conclusion that linguistic and indexical tasks may utilize two independent processes may be premature. When boys and girls are excluded from the analysis, the CI users resemble the normal hearing subjects listening to 8-channel simulations, as they did in the vowel identification task. Rather than concluding that two separate processes are involved,

these data may suggest that when the listener must make fine spectral distinctions, such as is required to distinguish talkers who share a similar range of vocal pitch, both CI users and normal hearing subjects listening to CI simulations perform comparably due to similar processes.

## The Present Study

Understanding how linguistic and indexical information interact in speech perception may provide new insight into possible training methodologies for newly implanted individuals. Given that there are no standardized training and rehabilitation protocols available to CI users, the source of the variability in benefit and outcome are further confounded with experience. Would listeners benefit from explicit training after implantation, and if so, what type of training is most appropriate? Given that most previous research has focused exclusively on linguistic tasks (Fu, Galvin, Wang & Nogaki, 2005), it is unknown whether training on nonlinguistic tasks will also promote robust generalization and transfer. Moreover, does the level of attention required to perform the training task modulate the amount of learning that is observed following training?

The present study compared how training on a linguistic versus indexical task affected listeners' ability to accurately perceive words in sentences. Using sentences processed with an 8-channel sinewave vocoder, normal hearing subjects were trained to identify either the gender or identity of six talkers, or transcribe their speech. Pre- to post-test comparisons of transcription accuracy scores assessed the effectiveness of training. Given the results of previous studies, we hypothesized that subjects trained on talker identification would perform better than those who were trained on gender identification. Moreover, we predicted that training on talker identification would match or exceed the performance of subjects trained on sentence transcription due to increased perceptual attention required to learn to identify the talkers from such severely spectrally degraded stimuli.

## Method

### Subjects

Seventy-eight normal-hearing young adults participated in the study (60 female, 18 male; mean age 21 years). All subjects were native speakers of American English. Most ( $n = 69$ ) were monolingual, with only nine reporting being fluent speakers of more than one language. Subjects were recruited from the Indiana University community, and either received monetary compensation for their participation (\$10 per session) or course credit in an Introductory Psychology class (1 credit per session). Of the seventy-eight subjects tested, six were excluded from the final data analysis (two failed to return for the generalization session, one failed to return in a timely manner, and three due to program errors). Of the 72 remaining subjects, 43 returned for the follow up portion of the experiment.

### Stimuli

Stimuli consisted of 212 meaningful (116 high predictability (HP), 48 low predictability (LP)), and 48 anomalous (AS) SPIN sentences (Kalikow, Stevens & Elliott, 1977; Clopper, Carter, Dillon, Hernandez, Pisoni, Clarke, Harnsberger & Herman, 2002). SPIN sentences are phonetically balanced for phoneme occurrence in English, and contain between five and eight words, the last of which is the keyword to be identified. In the HP sentences, the final word is highly constrained by the preceding semantic context (e.g., "A bicycle has two wheels."), whereas in the LP sentences the preceding context is uninformative (e.g., "The old man talked about the lungs."). The AS sentences retain the overall format of their meaningful counterparts, except that all words in the sentence are semantically unrelated,

resulting in a sentence that preserves proper syntactic structure, but is semantically anomalous (e.g., “The round lion held a flood.”). A passage of connected speech (Rainbow Passage; Fairbanks, 1940) was used during the familiarization portion of the experiment. Wavefiles of the materials were obtained from the Nationwide Speech Corpus (Clopper, 2004). Materials were produced by 8 speakers (4 male, 4 female) from the midland dialect.

## Synthesis

Stimulus processing was conducted in Tiger CIS (<http://www.tigerspeech.com/>) and simulated an 8-channel cochlear implant using the CIS processing strategy. Stimulus processing involved two phases, an analysis phase, which divided the signal into bands and derived the amplitude envelope from each band; and a synthesis phase, which replaced the frequency content of each band with a sinusoid that was modulated with its matched amplitude envelope. Analysis used band-pass filters to divide the stimuli into 8 spectral channels between 200 and 7000 Hz 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 and a 24 dB/octave slope. Subsets of the materials to be used in the generalization phase were processed with four and six channels, to further reduce the amount of information in the signal. All stimuli were saved as 22 kHz sampling rate 16-bit windows PCM wav files, and normalized to 65 dB RMS (Level v2.0.3, Tice & Carrell, 1998) to ensure that stimuli were equal in intensity across all materials, and that no peak clipping occurred.

## Procedures

All methods and materials were approved by the Human Subjects Committee and Institutional Review Board at Indiana University Bloomington. For data collection, a custom script was written for PsyScript and implemented on four Apple PowerMac G4 each with a 15-inch color LCD monitor. Audio signals were presented over Beyer Dynamic DT-100 headphones, calibrated with a voltmeter to a 1000 Hz tone at 70 dBv SPL. 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. Before the presentation of each audio signal, a fixation cross was presented at the center of the screen for 500 milliseconds to alert the subject to the upcoming trial. Following stimulus offset, the subject was prompted to make their response. A 1000 millisecond interval separated each trial. For the transcription trials, a dialog box was presented on the screen prompting subjects to type in what they heard. For talker identification, subjects clicked on the one box (out of six) that contained the name of the talker that produced the sentence. For gender identification, subjects clicked on a box labeled “female” or “male”. There were no time limits for responding, and subjects pressed a button to advance to the next trial. Subjects performed at their own pace, and were allowed to rest between blocks as needed. The experimental session lasted approximately 40-60 minutes.

**Training.** Training took place over two sessions. The materials and tasks varied across blocks, but the same block structure was used for all groups, and all stimuli were randomized within each block. Session 1 began with two pre-test blocks in order to establish a baseline level of performance before training (Table 1). In block 1, subjects transcribed 30 unique LP sentences, and 30 unique AS sentences in block 2. In these blocks, the subjects simply transcribed the sentences, and received no feedback.

In the familiarization phase (Block 3) subjects passively listened to the Rainbow passage produced by each of the six talkers in order to familiarize them with the voices and synthesis condition, and teach them the appropriate labels that would be used during training. Although subjects in all three training groups heard the same materials, they were required to make different responses during training.

During familiarization, subjects in the Talker ID group were presented with the passage paired with the name of the talker who produced it (Jeff, Max, Todd, Beth, Kim, Sue). Subjects were informed that they would be asked to identify the talkers by name, and to listen carefully for any information that would help them learn to recognize the talkers’ voice. Subjects in the Gender ID group heard the same passages, but paired with the appropriate gender label (Male or Female) for each talker. These subjects were informed that they would be asked to identify the gender of the talkers, and to listen carefully for any information that would help them learn to recognize each talker’s gender. Subjects in the Transcription group heard each passage presented along with the name of the talker who produced it (Jeff, Max, Todd, Beth, Kim, or Sue), but were informed that they would be asked to transcribe sentences produced by each talker, and to listen carefully in order to better understand the degraded signals.

Blocks 1-2	Block 3	Blocks 4-6
Pre-test	Familiarization	Training
Transcribe: 30 LP and 30 AS sentences	Passively listen: Rainbow passage	Transcribe, ID Talker, or ID Gender: 150 HP sentences

**Table 1.** Session 1 assessed the pre-test transcription abilities of subjects before training, familiarized them with the talkers and materials, and initiated training. The tasks that subjects performed and the materials that were presented in each block of Session 1 are listed in the table.

The training blocks (4, 5 and 6) consisted of 150 HP sentences. Each talker produced the same 25 sentences, so that subjects would hear six versions of each sentence in order to learn characteristics of the individual voices. During the training trials, subjects were presented with a sentence and asked to make a response appropriate for their training group. Subjects in the Talker ID group were asked identify the correct talker by clicking one of six buttons on the computer screen labeled with the talkers’ names. After the subject indicated their response, a red circle appeared around the name of the correct talker as feedback. Subjects in the Gender ID group responded by clicking one of two buttons on the computer screen that contained the appropriate gender label. After the subject indicated their response, a red circle appeared around the correct gender of the talker as feedback. Subjects in the Transcription training group were asked to type what they thought the talker said, and received the correct transcription of the sentence as feedback. For all training groups, feedback was provided regardless of the accuracy of the subject’s response.

Session 2 (Table 2) was completed within 3 days of session 1, and began with a repetition of the familiarization phase (block 7) in which subjects again heard the rainbow passage produced by each talker. The purpose of this block was to re-familiarize the listener with the voices and labels, since at least 24 hours had passed since the first training session. Two training blocks followed, consisting of 90 HP sentences. Again, subjects received feedback regardless of their performance.

Generalization and transfer of training were tested in blocks 10, 11 and 12, and subjects were asked to transcribe novel materials that they had not heard earlier during the experiment. In block 10, the transfer of training to more severe spectral degradation was assessed using 36 unique HP sentences, half of which were processed with 4-channel sinewave vocoder, and the other half with a 6-channel sinewave vocoder. Generalization of training to novel materials by familiar talkers was assessed in block 11 with 18 AS and 18 LP sentences processed with the same 8-channel vocoder used during training. In block 12,

transfer of perceptual learning to novel talkers was assessed using 20 unique HP sentences produced by two new talkers (1 male, 1 female). Following generalization, two post-test blocks (13 and 14) assessed the relative gains in performance due to training. In block 13 subjects transcribed a selection of twelve AS sentences from pre-test block 2, whereas in block 14 subjects transcribed a selection of twelve LP sentences selected from pre-test block 1.

Block 7	Blocks 8-9	Block 10	Block 11	Block 12	Blocks 13-14
Familiarization	Training	Generalization: More degraded	Generalization: Novel materials	Generalization: Novel talkers	Post-test
Passively listen: Rainbow passage	Transcribe, ID Talker, or ID Gender: 90 HP sentences	Transcribe: 18 HP (4-band) 18 HP (6-band) sentences	Transcribe: 18 AS & 18 LP sentences	Transcribe: 20 HP sentences	Transcribe: 12 AS & 12 LP sentences (from pre-test)

**Table 2.** Session 2 featured a continuation of training, followed by tests of generalization to new materials and the post-test (both transcription tasks). The tasks that subjects performed and the materials that were presented in each block of Session 2 are listed in the table.

**Retention.** One month after the initial training sessions, subjects returned for a third session to assess long-term retention of training (Table 3). During the retention test, subjects transcribed the same materials from generalization and post-test blocks 10 through 14. The purpose of this retention session was to assess how well perceptual learning was maintained over time, and to discern whether training differentially affected the long-term retention of training.

Block 15	Block 16	Block 17	Blocks 18-19
Generalization: More degraded	Generalization: Novel materials	Generalization: Novel talkers	Post-test
Transcribe: 18 HP (4-band) 18 HP (6-band) sentences	Transcribe: 18 AS & 18 LP sentences	Transcribe: 20 HP sentences	Transcribe: 12 AS & 12 LP sentences (from pre-test)

**Table 3.** Session 3 occurred 1 month after session 2, and tested subjects abilities to transcribe the materials that they experienced in session 2 to assess the stability of training over time. The tasks that subjects performed and the materials that were presented in each block of Session 3 are listed in the table.

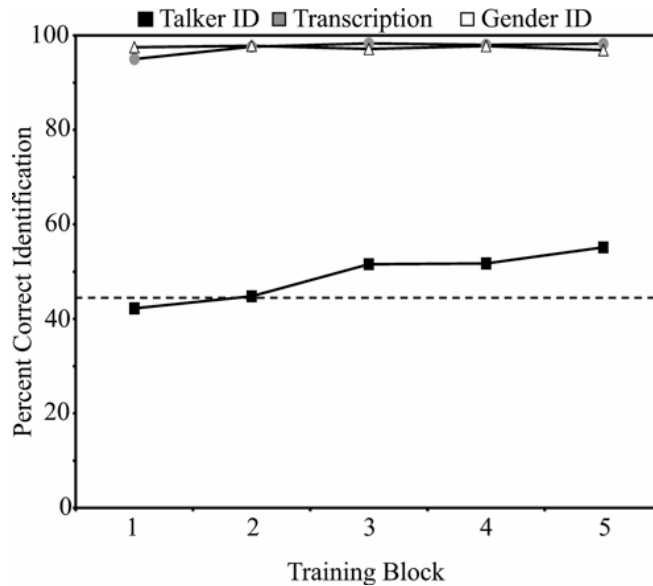
**Analysis and Scoring.** Keyword accuracy scores were based on the final word in each sentence. Common misspellings and homophones were counted as correct responses, but words with added or deleted morphemes were counted as incorrect. Perceptual learning during training was assessed by comparing performance across the five training blocks. Pre- to post-test comparisons provided an assessment of the relative gains from training across the three training groups. Comparison of performance at pre- and post-test to performance on new materials provided an assessment of generalization of training to novel stimuli. Generalization was said to have occurred if performance was significantly higher than the pre-test and greater than or equal to that at post-test. Comparison of pre- and

post-test performance to performance on new talkers provided an assessment of transfer of training to novel talkers. Comparisons of performance on the 4-band and 6-band stimuli provided an assessment of how well training transferred to more severely degraded stimuli. Comparison of performance in session 2 with performance in session 3 provided an estimate of long-term retention of training. A measurement of savings was calculated for each type of material by dividing performance in session 2 by that in session 3 and normalizing to one (e.g.,  $4\text{-band}_{\text{savings}} = 1 - (4\text{-band}_2/4\text{-band}_3)$ ). This provided an estimate of how robust perceptual learning was over time.

## Results

### Perceptual Learning during Training

Accuracy on the training tasks varied by training group (Figure 1). Subjects in the Gender ID and Transcription training groups performed near ceiling and subjects from the Talker ID group performed just above chance.



**FIGURE 1.** Perceptual learning across the five training blocks. The dashed horizontal line indicates the level of performance that subjects must exceed in order to be considered significantly different from chance in the talker identification condition. Subjects trained to transcribe the sentences (Transcription) appear as filled circles. Subjects trained to identify the gender of the talker (Gender ID) appear as filled triangles. Subjects trained to identify the talkers by their voices (Talker ID) appear as filled squares.

Subjects in the Transcription training group performed extremely well across all five training blocks. In block 1, subjects correctly identified 95% of the keywords and performance reached ceiling in block 2 (98% correct) and remained at ceiling for the last three training blocks. A univariate ANOVA revealed a significant main effect of Block ( $F(4, 190) = 6.441, p < 0.001$ ), indicating that subjects showed improvement across training blocks. Post hoc Bonferonni tests revealed that subject performance in block 1 was significantly lower than performance in all other blocks (all  $p < 0.009$ ). Performance in blocks 2 through 5 did not differ from one another (all  $p > 0.88$ ). A trend toward a main effect for Talker

Gender was observed ( $F(1, 190) = 3.156, p = 0.077$ ), with female speech being transcribed more accurately than male speech.

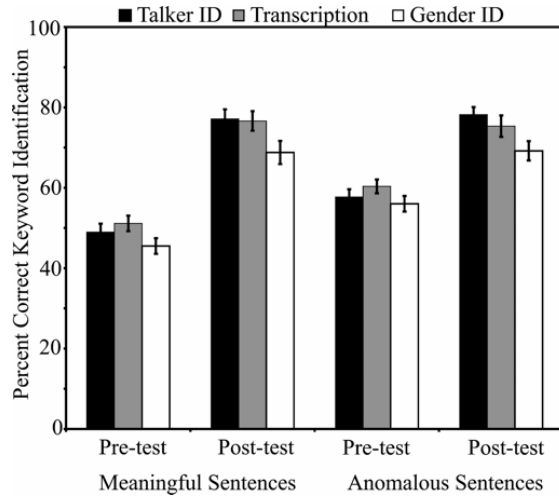
Subjects' accuracy in the Gender ID training condition was also extremely high across all five training blocks. Subjects' ability to identify the gender of the talkers was at ceiling (>95%) in all training blocks. Main effects for Block ( $F(4, 190) = .228, p = 0.922$ ), and Talker Gender ( $F(1, 190) = 1.324, p = 0.251$ ) were not observed, indicating that subject performance did not vary across blocks, and was equal for male and female talkers.

Performance of the Talker ID group was considerably more variable across subjects. Since inter-gender confusions (identifying male talkers as female, or female talkers as male) were rare, occurring less than 2 percent of the time, a more conservative level of chance was used (1 out of 3 rather than 1 out of 6). According to the binomial probability distribution, performance must be at least 44.46% correct to significantly exceed chance. Most subjects ( $n = 26$ ) were able to identify talkers at a level greater than chance beginning in block 2 and showed improvement as training progressed (Block 1: 42.2%, Block 2: 44.8%, Block 3: 51.6%, Block 4: 51.7%, Block 5: 55.1%). A univariate ANOVA revealed a significant main effect of Block ( $F(4, 250) = 9.428, p < 0.001$ ) with subject performance improving significantly between blocks 1 and 5 ( $p < 0.001$ ). A significant main effect of Talker Gender was also observed ( $F(1, 250) = 39.509, p < 0.001$ ), with subjects identifying female talkers (54%) more accurately than male talkers (44%).

### Performance after Training

**Pre- to Post-test Comparisons.** Overall, the type of training a subject received determined how well they performed at post-test; however, all subjects showed significant gains in sentence transcription accuracy due to training (Figure 2). For the subjects in the Transcription training group, performance increased from 51% correct for the meaningful sentences at pre-test to 77% correct at post-test. Similar gains were observed for the anomalous sentences increasing from 60% correct at pre-test to 75% at posttest. A univariate ANOVA revealed a significant main effect of Materials ( $F(3, 152) = 32.136, p < 0.001$ ), with subjects performing significantly better on anomalous than meaningful sentences at pre-test but not at posttest ( $p < 0.001$  and  $p = 0.977$ , respectively). This effect is likely due to exposure, since the anomalous sentence pre-test always came after the meaningful sentence pre-test. This difference of 9% is within the normal range of gains expected from merely being exposed to the stimuli without engaging in explicit training, as documented by Davis and colleagues (2005). Furthermore, a significant main effect of Talker Gender ( $F(1, 152) = 5.939, p = 0.016$ ) was observed. Subjects were significantly more accurate at transcribing the speech of female talkers than male talkers.

Training on Gender identification successfully transferred to sentence transcription (Figure 2). For meaningful sentences, performance increased from 45% at pre-test to 69% correct at post-test. Similar gains were observed for the anomalous sentences increasing from 56% correct at pre-test to 69% at posttest. An ANOVA revealed a significant main effect of Materials ( $F(3, 152) = 25.959, p < 0.001$ ), indicating that performance varied according to the type of materials subjects were asked to transcribe. At pre-test, subjects performed significantly better on the anomalous sentences than the meaningful sentences ( $p = 0.006$ ), but were identical at posttest ( $p = 1.00$ ). A significant main effect was observed for Talker Gender ( $F(1, 152) = 10.222, p = 0.002$ ), again indicating that subjects were significantly more accurate at transcribing the speech of female talkers than male talkers.



**FIGURE 2.** Percent correct keyword identification scores for subjects trained on talker identification (Talker ID), gender identification (Gender ID) or sentence transcription (Transcription) on the pre- and post-test materials.

For subjects in the Talker ID group, a significant main effect of Materials was also observed ( $F(3, 200) = 69.555, p < 0.001$ ). For meaningful sentences (Figure 2), subjects improved significantly from pre- (48% correct) to post-test (75% correct,  $p < 0.001$ ). Similar findings were observed for the anomalous sentences, with performance increasing significantly from 56% correct to 79% correct ( $p < 0.001$ ). As was observed for subjects in the Transcription and Gender ID groups, performance was significantly better on anomalous sentences than meaningful sentences at pre-test ( $p = 0.003$ ), but identical at post-test ( $p = 0.652$ ). A significant main effect of Talker Gender was also observed ( $F(1, 200) = 72.664, p < 0.001$ ) indicating that the materials produced by female talkers were correctly transcribed significantly more accurately than those produced by male talkers.

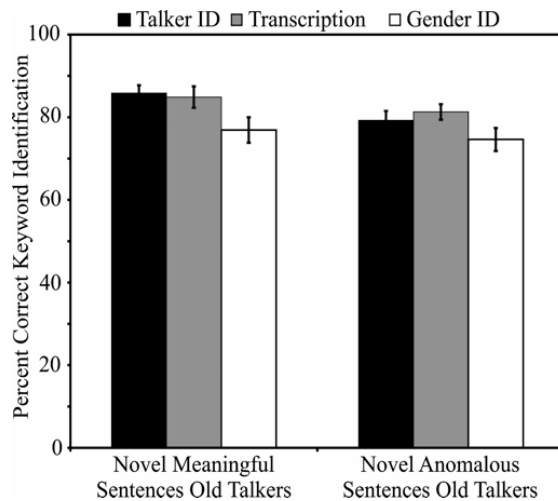
Univariate ANOVAs comparing the scores of all three training groups revealed that pre-test performance did not differ across training groups for the anomalous ( $F(2, 126) = 1.356, p = 0.262$ ) or meaningful sentences ( $F(2, 123) = 2.569, p = 0.081$ ) indicating that subjects in all groups performed at a comparable level before training began. Differences in performance emerged at post-test, for both the meaningful ( $F(2, 126) = 3.656, p = 0.029$ ) and anomalous sentences ( $F(2, 126) = 4.234, p = 0.017$ ). In both cases, subjects in the Gender ID training group performed less accurately than subjects in the Talker ID training ( $p = 0.036, p = 0.013$ ) and Transcription training groups ( $p = 0.075, p = 0.156$ ).

**Generalization to New Materials.** Overall, training successfully generalized to the transcription of novel sentences produced by familiar talkers (Figure 3). Transcription training successfully generalized to new meaningful sentences produced by the familiar talkers (85.7%). A univariate ANOVA revealed that there was a significant main effect of Session ( $F(3, 126) = 96.629, p < 0.001$ ), and Bonferonni tests indicated that subjects performance was significantly better for novel meaningful materials than at pre-test ( $p < 0.001$ ) or post-test ( $p = 0.014$ ). A similar finding was observed for the new anomalous sentences ( $F(2, 114) = 25.974, p < 0.001$ ), and subjects performed significantly better on the novel anomalous sentences (79.1%) than at pre-test ( $p < 0.001$ ) but not at post-test ( $p = 0.175$ ). Additionally, a significant main effect of Talker Gender was observed for both meaningful ( $F(1, 126) = 6.741, p = 0.010$ ) and anomalous sentences ( $F(1, 114) = 5.462, p = 0.021$ ), indicating that female talkers were again transcribed more accurately than male talkers.

Subjects trained to identify talker gender showed robust generalization to new meaningful (76.7%;  $F(3, 126) = 55.096, p < 0.001$ ) and anomalous sentences (74.4%;  $F(2, 114) = 17.593, p < 0.001$ ). For both anomalous and meaningful sentences, performance on the new materials was significantly higher than pre-test (both  $p < 0.001$ ) and did not differ from post-test (both  $p > 0.09$ ). A significant main effect of Talker Gender was observed for the new meaningful sentences ( $F(1, 126) = 10.058, p = 0.002$ ), but not new anomalous sentences ( $F(1, 114) = 2.746, p = 0.10$ ).

Subjects trained on Talker ID also showed robust generalization to new meaningful (84.7%;  $F(3, 200) = 136.095, p < 0.001$ ) and anomalous sentences (81.1%;  $F(2, 150) = 58.199, p < 0.001$ ). For both training groups, performance on the new materials was significantly more accurate than pretest (all  $p < 0.001$ ) and was greater than (meaningful sentences  $p < 0.001$ ) or equal to (anomalous sentences  $p = 1.00$ ) performance at post-test. A significant main effect of Talker Gender was observed for both new meaningful sentences ( $F(1, 200) = 38.217, p < 0.001$ ) and anomalous sentences ( $F(1, 150) = 30.201, p < 0.001$ ), and subjects were more accurate in transcribing the female talkers than the male talkers.

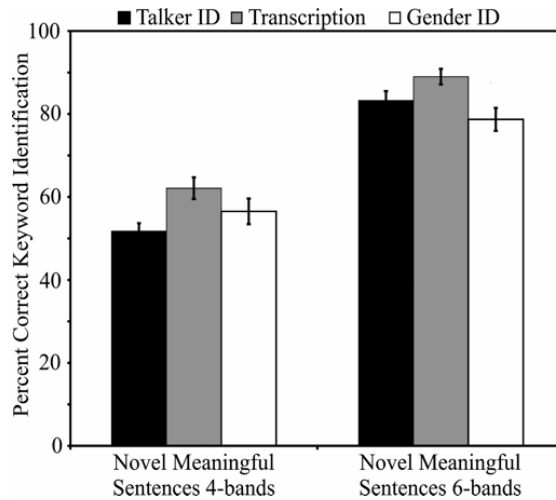
Comparison of performance on the meaningful sentences across all three groups using a univariate ANOVA revealed a significant main effect of Training ( $F(2, 126) = 6.403, p = 0.002$ ). Subjects in the Transcription group performed nearly identically to subjects in the Talker ID group ( $p = 0.932$ ), and both groups performed significantly better than the subjects in the Gender ID group ( $p = 0.015$  and  $p = 0.003$  respectively). In addition, a trend toward a significant main effect of Talker Gender was observed ( $F(1, 126) = 3.724, p = 0.056$ ) indicating that female talkers were transcribed with more accuracy than male talkers. Although a main effect of Training was not observed for the novel anomalous sentences ( $F(2, 126) = 2.795, p = 0.067$ ), a trend was observed for subjects in the Transcription training group to perform better than subjects in the Gender ID training group ( $p = 0.073$ ). A significant main effect of Talker Gender was also observed ( $F(1, 126) = 18.769, p < 0.001$ ) with subjects transcribing female talkers more accurately than male talkers.



**FIGURE 3.** Percent correct keyword identification scores on the new anomalous and meaningful sentences produced by familiar talkers (session 2, block 11) for subjects trained on talker identification (Talker ID), gender identification (Gender ID) or sentence transcription (Transcription).

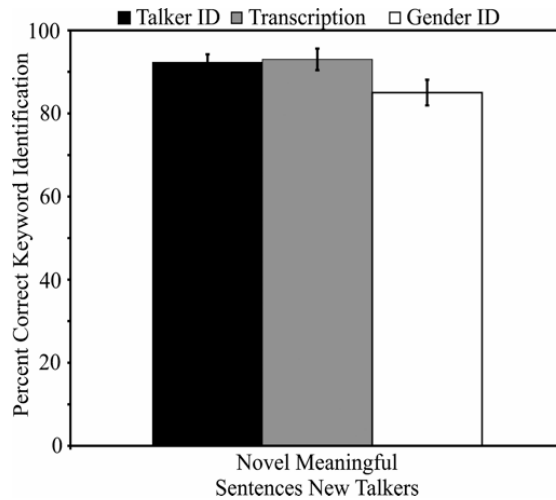
**Transfer of Training to Increased Spectral Degradation.** Subjects showed a graded response to stimuli that were more severe spectral degraded (Figure 4). Overall, subjects were more accurate at transcribing sentences in the 6-band processing condition (Transcription: 83.1%; Gender ID: 78.6%; Talker ID: 88.9%) than sentences in the 4-band processing condition (Transcription: 51.7%; Gender ID: 56.4%; Talker ID: 61.9%). A univariate ANOVA revealed a significant main effect of Processing for all groups (Transcription ( $F(1, 76) = 69.104, p < 0.001$ ); Gender ID ( $F(1, 76) = 29.731, p < 0.001$ ); Talker ID ( $F(1, 100) = 120.846, p < 0.001$ )), indicating that subjects performed significantly better on the 6-band sentences than the 4-band sentences. The main effect of Talker Gender was not significant for the Transcription training group ( $F(1, 76) = .066, p = 0.798$ ), or the Gender ID training group ( $F(1, 76) = 2.248, p = 0.138$ ), indicating that subjects performed equally well on male and female speech. Subjects in the Talker ID training group, however, did show a significant main effect of Talker Gender ( $F(1, 100) = 9.094, p = 0.003$ ), indicating that they transcribed the speech of female talkers more accurately than male talkers.

Comparison of the performance on the 4-band processed sentences across training groups using a univariate ANOVA revealed a significant main effect of Training ( $F(2, 126) = 4.44, p = 0.014$ ). Subjects in the Transcription training group performed significantly better than subjects in the Talker ID group ( $p = 0.01$ ), but did not differ from talkers in the Gender ID group ( $p = 0.399$ ). Subjects in the Talker ID training group performed similarly to subjects in the Gender ID group ( $p = 0.359$ ). The main effect of Talker Gender was not significant ( $F(1, 126) = .933, p = 0.336$ ). Comparison of performance on the 6-band stimuli across training groups also revealed significant main effect of Training ( $F(2, 126) = 4.702, p = 0.001$ ). Subjects in the Transcription group performed as well as subjects in the Talker ID group ( $p = 0.465$ ), but significantly better than subjects in the Gender ID group ( $p = 0.008$ ). Subjects in the Gender ID group performed as well as subjects in the Talker ID group ( $p = 0.213$ ). A significant main effect of Talker Gender was observed ( $F(1, 126) = 8.273, p = 0.005$ ), and subjects were significantly more accurate at transcribing the speech of female talkers than male talkers.



**FIGURE 4.** Percent correct keyword identification scores for subjects trained on talker identification (Talker ID), gender identification (Gender ID) or sentence transcription (Transcription) on the meaningful sentences produced by familiar talkers but processed to have more severe spectral degradation (Block 10).

**Transfer of Training to Novel Talkers.** Transcription of novel sentences produced by unfamiliar talkers was equivalent to or better than transcription of meaningful sentences produced by familiar talkers (Transcription 92.3% correct; Gender ID 85% correct, Talker ID 93% correct). For all training groups, performance on new talkers was significantly higher than pre-test and post-test (both  $p < 0.001$ ) suggesting that talker familiarity may not necessarily enhance transcription accuracy on CI simulations as compared to other types of spectral degradation (e.g. noise). Moreover, training-induced differences in performance were also observed (Figure 5), and a significant main effect of Training was again noted ( $F(2, 126) = 6.874, p < 0.001$ ). Subjects from the Transcription and Talker ID training groups performed the same ( $p = 0.951$ ), and significantly better than subjects in the Gender ID group ( $p = 0.004$  and  $p = 0.005$ , respectively).



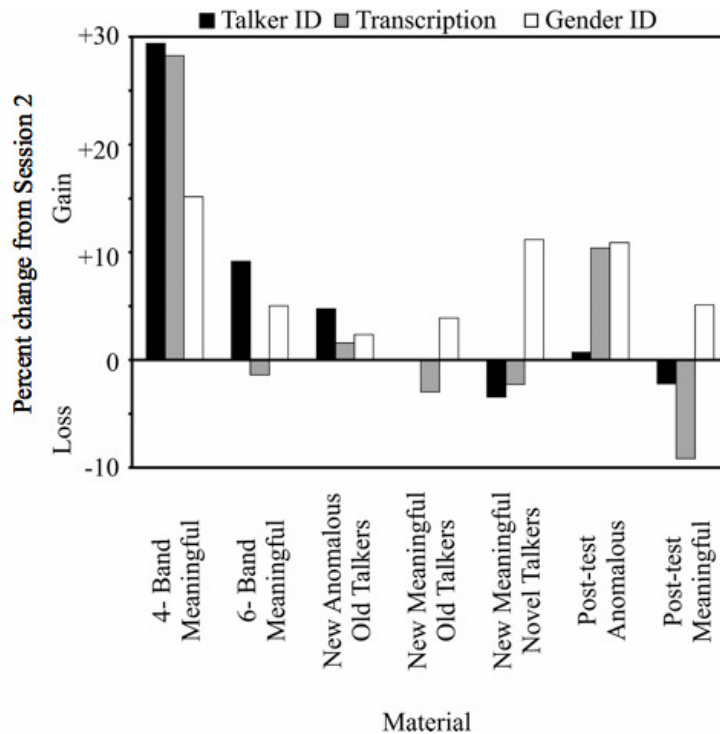
**FIGURE 5.** Percent correct keyword identification scores for subjects trained on talker identification (Talker ID), gender identification (Gender ID) or sentence transcription (Transcription) on the meaningful sentences produced by novel talkers (block 12).

**Retention of Training.**

Of the 72 subjects who participated in sessions 1 and 2, 43 returned for retention testing in session 3. Since fewer subjects overall participated in session 3, data were matched such that the analyses only compared the performance of subjects who attended all three sessions. A one-way ANOVA comparing performance in session 2 with that in session 3 revealed all subjects in the Transcription training group improved their performance on the 4-band stimuli ( $F(1, 34) = 9.092, p = 0.015$ ) from 57% in session 2, to 73% in session 3. In session 3, performance on all other materials (6-band, new anomalous sentences, new meaningful sentences, new talkers, post-test anomalous, post-test meaningful) did not change from session 2 (all  $p > 0.1$ ). Subjects in the Gender ID training group also showed significant gains on the 4-band stimuli in session 3 ( $F(1, 46) = 4.713, p = 0.035$ ), improving from 61% in session 2 to 70% in session 3. Improvements were also observed for the new meaningful sentences ( $F(1, 46) = 6.595, p = 0.014$ ), which increased from 79% in session 2 to 87% in session 3. Performance on all other materials (6-band, new anomalous sentences, new talkers, post-test anomalous, post-test meaningful) did not change from session 2 (all  $p > 0.1$ ). Subjects in the Talker ID group showed significant improvement on the 4- (51 to 66%, ( $F(1, 58) = 14.236, p < 0.001$ ) and 6-band stimuli (85 to 93%, ( $F(1, 58) = 5.353, p = 0.024$ )). Performance on all other materials (new anomalous sentences, new

meaningful sentences, new talkers, post-test anomalous, post-test meaningful) did not change from session 2 to session 3 (all  $p > 0.1$ ).

It is important to note that these retention tests included the same materials that appeared in the post-tests and generalization tests, so these measures are purely designed to show whether training is stable over time rather than to assess generalization to novel materials or conditions. To this end, a measure of savings was employed that divided the performance in session 2 by the performance in Session 3 and subtracting the result from one (e.g.,  $1 - (\text{Post-test}_3/\text{Post-test}_2)$ ) in order to determine the percent gain or loss that subjects received for each type of material (Figure 6). Across all materials, subjects in the Gender ID group showed the largest gain from Session 2 to Session 3 (increasing overall by 54%) followed by subjects in the Talker ID group (38%) and subjects in the Transcription training group (12%). The largest gains for all groups were observed for the 4-band vocoded stimuli, demonstrating that previous exposure to the more severely spectrally degraded materials tended to improve performance most at retention.



**FIGURE 6.** Percent gain or loss across groups as a function of testing materials. The amount of savings was calculated by dividing performance in Session 2 by performance in Session 3 and subtracting one from the result.

**Talker ID Training: Subgroups.**

An additional finding of the present study emerged when first assessing subject performance on the Talker ID training task. As noted earlier, most ( $n = 26$ ) subjects could be trained to successfully identify talkers at a level greater than chance (44.3%). There was an additional subset of subjects, however, who could not, and were excluded from the analysis for the Talker ID group. Unlike the good

learners, these poor learners ( $n = 5$ ) were never able to identify talkers at a level greater than chance in any of the blocks (Block 1: 30.8%, Block 2: 35.4%, Block 3: 36.3%, Block 4: 34.7%, Block 5: 32.9%), as indicated by a univariate ANOVA ( $F(4, 40) = 0.05, p = 0.628$ ). A significant main effect of Talker Gender was found, however ( $F(1, 40) = 9.941, p = 0.003$ ), revealing that female talkers were correctly identified significantly more often (38%) than male talkers (30%), as was the case in the good learners.

Furthermore, subjects who could not identify the talkers at a level exceeding chance performed significantly more poorly on the transcription tasks than the subjects who were proficient at talker identification. A series of one-way ANOVAs revealed that performance did not differ at pre-test for either the meaningful ( $p = 0.105$ ) or anomalous sentences ( $p = 0.310$ ). After training, however, a significant main effect of Group was observed for all materials (all  $p < 0.003$ ), indicating that although subjects performed the same at pre-test, their performance increased at a different rate depending on how well they could perform the training task. Such a result is not likely to be caused by inattention, or laziness on the part of the participants in the poor learning group, since the transcription errors they made were phonologically related to the target words, and response omissions were no more prevalent than in the good learning group. Rather, it appears that the ability to detect and utilize acoustic information important for the indexical training task is related to the ability to extract acoustic information important for recognizing the linguistic content of utterances.

## Discussion

The present study compared training that selectively focuses the listener's attention on the indexical information in the speech signal to training that focuses entirely on the linguistic content. Although all three types of training in this experiment produced significant pre- to post-test gains in performance, talker identification and sentence transcription training appeared to provide the largest and most robust overall improvement (Figure 2). Generalization to new materials and talkers was equivalent for the talker identification and transcription trained subjects, both of whom performed better than the subjects trained on gender identification (Figure 3 and Figure 5). Generalization to materials that were more spectrally degraded showed a mixed pattern of results (Figure 4). For stimuli that were more severely spectrally degraded (4- and 6- band), subjects trained on sentence transcription performed best, subjects trained on gender identification performed least accurately and subjects trained on talker identification displayed an intermediate level of performance. No effect of talker familiarity was observed. Subjects performed as well or better on the new talkers than they did on the old talkers, suggesting that the benefit of talker familiarity may not be as robust under cochlear implant simulations as compared to other forms of degradation (e.g., noise). However, baseline intelligibility for these talkers has not been established so it is possible that the talkers in the "new talker" condition were intrinsically more intelligible than the "old talkers" used in the training blocks.

Two main conclusions can be drawn from these data. The first is that training on an indexical task yields equivalent results to traditional linguistic training using transcription tasks if the task demands are high enough to require sustained attention. Evidence for this comes from the across group comparisons of post-test and generalization scores for the subjects in the Talker ID group, who performed similarly to the subjects in the Transcription training group, but significantly better than the subjects in the Gender ID training group (Figure 2). Compared to gender identification (which was at ceiling in the first training block), talker identification training is a difficult task under cochlear implant simulations, requiring high levels of attention and focus. Moreover, when a listener is exposed to a speech signal that is meaningful in their native language they cannot help but to process it as such. Even though subjects' attention in the Talker and Gender ID tasks were not directed toward the linguistic

information in the signal, presumably they still processed the linguistic content of the sentences automatically.

The second main finding is that the benefit of exposure may be determined by whether the subject can successfully access the acoustic information in the speech signal. Subjects in the Talker ID group, who had to make fine acoustic distinctions among voices, performed significantly better than subjects in the Gender ID group. Moreover, the subjects from the Talker ID group who could not learn to identify the talkers at a level greater than chance performed significantly worse on sentence transcription than subjects who could identify the talkers. Taken together, these findings suggest that the access and attention to fine acoustic details learned during talker identification training may enhance a listener's ability to extract linguistic information.

### **Differences in Task Demands and Attentional Resources**

The data from the present study suggest that interactions between attentional demands and task difficulty may play a large role in determining the amount of benefit that a subject will receive from training. Talker identification under a CI simulation is considerably more difficult than under normal acoustic conditions. The acoustic information that specifies the voice of the talker in the natural signal appears to be significantly degraded when processed through a cochlear implant speech processor, whereas the acoustic information needed to successfully identify the gender of a talker under 8 channel CI-simulation is relatively intact. Thus, the task demands placed on a listener are significantly higher in a talker identification task than those in a gender identification task. Subjects in the Talker ID training group, while performing significantly greater than chance, only achieved an average score of 55% correct talker recognition on the final day of training; subjects in the gender ID training group were at ceiling from the first training block. These results suggest that the identifying characteristics of a talker's voice may rely on detailed spectral cues within specific frequency regions. Such cues are not well preserved in a cochlear implant. Gender identity cues, on the other hand, may rely more on spectral information across a wider range of frequencies and the relative spectral weighting of information in each frequency band in the vocoder may allow listeners to perform more accurately.

The differences in the availability of acoustic information may have produced differences in task demands. More attention is required when making fine-grained distinctions between talkers' voices, and comparably less is required to distinguish genders. These differences in attentional requirements may explain the differences in post-test gains and strength of generalization. Subjects who were required to perform a more demanding task during training performed better in the post-test and generalization phase than subjects who performed a less demanding task. Additionally, talker identification may require the utilization of cues from many different aspects of phonological structure (i.e., prosody, stress patterns, speaking rate, etc.), which are apparent in longer speech samples and require sustained attention for a longer period of time as compared to cues for gender identity. After the experiment, some subjects in the talker identification group said that they focused their attention on distinctions in overall speaking patterns and pronunciation habits in order to distinguish the talkers. As such, listening attentively to longer samples of speech may have resulted the perception of more of the linguistic information in the signal. If subjects in the gender identification group could make a decision more rapidly based on lower level acoustic cues, they may not have attended to the signal as long, and may not have received as much of a benefit from the mandatory linguistic processing.

One might expect subjects in the Talker and Gender ID training conditions to perform worse on the post-test and generalization tests than subjects in the Transcription training condition due to subjects performing fundamentally different tasks than they were trained on. Subjects in the Transcription group,

however, performed similarly to those in the Talker ID group, suggesting that training on two different tasks can produce an equivalent benefit. For subjects in the talker identification group, perceptual learning transferred from training to testing even though they were performing a different task in each condition. Subjects in the Transcription group help to establish what levels of generalization should be expected, since they performed the same task during both training and testing. Subjects performed similarly in the talker identification condition, but significantly more poorly in the gender identification condition. This finding suggests that additional attentional demands during training may help to overcome the differences in the tasks.

Although the differences in performance were only observed in the short term, the equivalence of performance across the three groups at the retention session could simply be a factor of the familiarity with the materials. Subjects were tested on the same materials that they were exposed to during the first testing session rather than on novel materials from the same talkers. It could be the case that performance on a true generalization and retention test consisting of completely novel materials may distinguish group performance across the different training conditions. Due to the limitation of available stimulus materials, however, this could not be assessed by the present experiment.

### **Access to the Acoustic Information in the Signal**

It is important to note that not all subjects were able to learn the talkers' voices over the five training blocks. Although the vast majority of subjects (84%) could learn to identify the talkers at a level greater than chance, several subjects could not. Although transcription scores at pre-test were comparable for both groups, the subjects who could not learn to identify the talkers by voice performed significantly worse on sentence transcription in the post-test and generalization blocks. Moreover, these differences could not be attributed to inattention or disinterest, since transcription errors were largely phonologically relevant and demographic variables such hearing insult or speech pathology problems did not reveal any abnormalities.

Additionally, previous research supports the proposal that the ability to learn to identify talkers by voice can predict transcription accuracy for speech samples produced by these talkers. In the original study demonstrating the transfer of talker identification training to word identification accuracy, Nygaard, Sommers, and Pisoni (1994) reported that not all of their subjects were able to learn to identify talkers by voice. Subjects who could successfully identify the talkers by voice showed higher recognition accuracy scores for words produced by familiar talkers as compared to novel talkers. The subjects who could not learn to identify the talkers by voice did not show such a difference. Taken together, the present finding suggest that it is not the mere exposure to a talker or a synthesis condition that is responsible for the gains observed after training, but rather the ability to access and utilize the acoustic information required to recognize the talkers by voice.

The findings of the present study also replicate those of Cleary and colleagues (2005) who examined talker discrimination in a group of pediatric cochlear implant users. Children listened to pairs of sentences and decided whether the two sentences were produced by the same or different talkers. Considerable variability was observed among the children with CIs, but those who were more proficient at talker discrimination also showed increased accuracy on a word identification task (Cleary et al., 2005). Taken together with the findings of Cleary and colleagues and Nygaard and colleagues, these data provide strong evidence for the interaction of lexical and indexical information, and suggest that the two streams may indeed be encoded and processed together.

There was no clear effect of talker familiarity on the recognition of speech processed by a CI simulation; subjects were as accurate at transcribing the speech produced by novel talkers as they were at transcribing speech produced by talkers used during training. The lack of a talker familiarity effect using CI simulated speech may not be completely anomalous, however. Barker (2006) showed a similar pattern of results for adult CI users trained to identify the voices of six talkers. In her study, CI users showed no differences in transcription accuracy performance for familiar versus unfamiliar talkers at a signal to noise ratio of +10 dB SNR. At 59% correct, talker ID accuracy scores for her fifteen CI users were nearly identical to our results with normal hearing subjects listening to 8-channel sinewave vocoders. Although she used a control group of normal hearing subjects, they performed the talker identification training with the unprocessed speech stimuli, so a direct comparison is inappropriate. Taken together, these data suggest that although indexical information regarding talker identity is preserved in electric hearing (as well as in acoustic simulations thereof), the talker familiarity effects that are observed for natural speech may differ in fundamental ways from those for cochlear implant simulations or individuals with CIs.

### **Behavioral and Clinical Implications**

The findings from the present study suggest that there are multiple routes to the perceptual learning of speech. Although most studies utilize traditional methods of training that exclusively focus the listener's attention on the symbolic linguistic content encoded in the speech signal (e.g., Fu et al., 2005), other routes can yield similar outcomes and benefits. The crucial factor seems to be the amount of attention that is required of the subject, and the degree to which performance can be improved. Tasks that require significant amounts of controlled attention to the indexical properties of the signal can be just as effective as tasks that rely exclusively on attention to the linguistic content of the message. This finding has important implications for training and rehabilitation strategies for individuals who receive cochlear implants. The benefit observed in the current study for non-traditional training methods suggests that a variety of stimulus materials could be utilized to maximize outcome. Instruction on how to auditorily distinguish individual voices may provide the CI user with a more stable foundation for voice recognition that can be generalized to new talkers in new situations. Additionally, including a variety of stimulus materials and challenging perceptual tasks may promote interest in training, and protect against boredom and fatigue that can occur when only a single task is used.

Although the overall goal of cochlear implantation has been to restore receptive auditory capacity to the severely hearing-impaired individual, there are many other nonlinguistic aspects to hearing on which a CI user could experience benefit. Sound localization, the detection and identification of environmental stimuli and the enjoyment of music are all aspects of normal hearing that have not been well investigated in cochlear implant populations. Since all of these tasks require attention to acoustic information encoded in the signal that is nonlinguistic, greater variety in training tasks and materials may yield more robust results, many of which may transfer to speech perception and language processing tasks. If the goal of cochlear implantation is to provide the listener with access to the acoustic world, we should begin focusing training on achieving on such a goal. By limiting training to linguistic tasks, we may be undermining the robust adaptive abilities of CI users by depriving them of the full benefit that they may one day enjoy. Speech is not isolated from the rest of the acoustic world in which we live. A decision needs to be made as to whether the goal of cochlear implantation is only to provide access to the speech signal or to replace hearing, and directed measures need to be taken to achieve these goals accordingly.

## References

- Barker, B.A. (2006). An examination of the effect of talker familiarity on the sentence recognition skills of CI users. Unpublished Doctoral Dissertation, University of Iowa.
- Bond, Z.S. & Moore, T.J. (1994). A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Communication, 14*(4), 325-337.
- Bradlow, A.R. & Bent, T. (In Press). Perceptual adaptation to non-native speech. *Cognition*.
- Bradlow, A.R., Torretta, G.M. & 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. Unpublished Doctoral Dissertation, Indiana University, Bloomington.
- Clark, G.M. (2002). Learning to understand speech with the cochlear implant. In Fahle, M and Poggio, T. (Eds), *Perceptual Learning*, Pp. 147-160. Cambridge: MIT Press.
- Clarke, C.M. & Garrett, M.F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America, 116*(6), 3647-3658.
- Cleary, M., Pisoni, D.B. & Kirk, K.I. (2005). Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. *Journal of Speech, Language, and Hearing Research, 48*, 204-223.
- Cleary, M. & Pisoni, D.B. (2002). Talker discrimination by prelingually deaf children with cochlear implants: Preliminary results. *Annals of Otolaryngology, Rhinoogy and Laryngology, 111*(5-2, Supplement. 189), 113-118.
- Clopper, C.G., Carter, A.K., Dillon, C.M., Hernandez, L.R., Pisoni, D.B., Clarke, C.M., Harnsberger, J.D., and Herman, R. (2001), The Indiana Speech Project: An overview of the development of a multi-talker multi-dialect speech corpus. In *Research on Spoken Language Processing Progress Report No. 25* (pp. 367-380). Bloomington, IN: Speech Research Laboratory, Indiana University.
- Clopper, C.G. (2004). Linguistic experience and the perceptual classification of dialect variation. Unpublished Doctoral Dissertation, Indiana University, Bloomington.
- Cox, R.M., Alexander, G.C. & Gilmore, C. (1987). Development of the Connected Speech Test (CST). *Ear and Hearing, 8*(5) Supplement, 119s.
- Davis, M.H., Johnsrude, I.S., Hervais-Adelman, A., Taylor, K. & McGettigan, C. (2005) Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General, 134*(2), 222-241.
- Dorman, M.F., Loizou, P.C. & 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. & 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.
- Dorman, M., Loizou, P., Fitzke, J. & Tu, Z. (1998). The recognition of sentences in noise by normal hearing listeners using simulations of cochlear implant signal processors with 6-20 channels. *Journal of the Acoustical Society of America, 104*(6), 3583-3585.
- Dupoux, E. & Green, K.P. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 914-927.
- Eisner, F. & McQueen, J.M. (2005). The specificity of perceptual learning in speech processing. *Perception and Psychophysics, 67*(2), 224-238.
- Fairbanks, G. (1940). *Voice and Articulation Drillbook*. New York: Harper and Row.
- Fahle, M. & Poggio, T. (Eds.). (2002). *Perceptual Learning*. Cambridge: MIT Press.
- Fu, Q-J., Chinchilla, S. & Galvin, J.J. (2004). The role of spectral and temporal cues on voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology, 5*, 253-260.
- Fu, Q-J., Chinchilla, S., Nogaki, G. & Galvin, J.J. (2005). Voice gender identification by cochlear implant users: The role of spectral and temporal resolution. *Journal of the Acoustical Society of America, 118*, 1711-1718.
- Fu, Q-J., Galvin, J.J., Wang, X. & Nogaki, G., (2005). Moderate auditory training can improve speech performance of adult cochlear implant patients. *Acoustic Research Letters Online, 6*(3), 106-111.
- Goldstone, R.L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585-612.

- Gonzales, J. & 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(1), 461-470.
- Greenspan, S.L., Nusbaum, H.C. & Pisoni, D.B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, 421-433.
- Hood, J.D. & Poole, J.P. (1980). Influence of the speaker and other factors affecting speech intelligibility. *Audiology*, 19(5), 434-55.
- Kalikow, D.N., Stevens, K.N. & Elliot, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, 61, 1337-1351.
- Krajlic, T. & Samuel, A.S. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin Review*, 13(2), 262-268.
- Ladefoged, P. & Broadbent, D.E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, 29(1), 98-104.
- Loebach, J.L. & Pisoni, D.B. (Under Review). Perceptual learning of spectrally degraded speech. *Journal of the Acoustical Society of America*.
- Luo, X., and Fu, Q. J. (2005). Speaker normalization for Chinese vowel recognition in cochlear implants. *IEEE Transactions on Biomedical Engineering*, 52, 1358-1361.
- McGarr, N.S. (1983). The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech and Hearing Research*, 26, 451-458.
- National Institutes of Health. (1995). Cochlear implants in adults and children. *NIH Consensus statement*, 13, 1-29.
- Nygaard, L.C., Sommers, M.S. & Pisoni, D.B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5(1), 42-46.
- Nygaard, L.C. & Pisoni, D.B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics*, 60(3), 355-376.
- Pallier, C., Sebastian-Gallés, N., Dupoux, E., Christophe, A. & Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory and Cognition*, 26(4), 844-851.
- Remez, R.E., Fellowes, J.M. & Rubin, P.E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, 23(3), 651-666.
- Schwab, E.C., Nusbaum, H.C. & Pisoni, D.B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors*, 27, 395-408.
- Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J. & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 303-304.
- Shannon, R.V., Fu, Q.-J. & Galvin, J. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Otolaryngocia Supplementum*, 552, 1-5.
- Shannon, R.V. (2005). Speech and music have different requirements for spectral resolution. *International Review of Neurobiology*, 70, 121-134.
- Sheffert, S.M., Pisoni, D.B., Fellowes, J.M. & Remez, R.E. (2002). Learning to recognize talkers from natural, sinewave and reversed speech samples. *Journal of Experimental Psychology*, 28(6), 1447-1469.
- Vongphoe, M. & Zeng, F.G. (2005). Speaker recognition with temporal cues in acoustic and electric hearing. *Journal of the Acoustical Society of America*, 118, 1055-1061.
- Weil, S.A. (2001). Foreign accented speech: Adaptation and generalization. Unpublished Master's Thesis, Ohio State University.