# Multiple routes to the perceptual learning of speech

Jeremy L. Loebach[a] and Tessa Bent[b]
*Speech Research Laboratory, Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana 47405 and DeVault Otologic Research Laboratory, Department of Otolaryngology-Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, Indiana 46202*

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

(Received 21 August 2007; revised 6 April 2008; accepted 18 April 2008)

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 and indexical tasks when adapting to cochlear implant simulations. Listening to sentences processed with an eight-channel sinewave vocoder, three separate 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 posttest comparisons demonstrated that training produced significant improvement for all groups. Moreover, subjects from the talker ID and transcription training groups performed similarly at posttest and generalization, and significantly better than the subjects from the gender ID training group. These results suggest that training on an indexical task that requires high levels of controlled attention can provide equivalent benefits to training on a linguistic task. When listeners selectively focus their attention on the extralinguistic information in the speech signal, they still extract linguistic information, the degree to which they do so, however, appears to be task dependent. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2931948]

PACS number(s): 43.71.Bp, 43.71.Es, 43.66.Ts, 43.66.Lj [MSS] Pages: 552–561

## I. 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 (such as gender, age and dialect, Abercrombie, 1967; Ladefoged and Broadbent, 1957). How these two types of information interact during speech perception is still poorly understood. 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 the perceptual learning of spectrally degraded speech. Using sentences that had been processed by a cochlear implant (CI) simulator, we investigated how different types of training affected perceptual learning and generalization to new sentences and sentences under more severely spectrally degraded conditions. We found that the amount of controlled attention required during the training task modulated the relative strength of perceptual learning. Training on talker identification, an indexical task that required a higher degree of attentional control and focus on the acoustic information in the signal, was just as effective as transcription training and both tasks elicited more robust generalization than training on gender identification.

### A. Indexical information and cochlear implants

Research with CI users has focused almost exclusively on speech perception, leaving the perception of other types of information carried in the acoustic signal (e.g., indexical information) unexplored. For linguistic tasks, acoustic simulations of cochlear implants have provided a useful tool for determining what acoustic information is necessary for accurate speech perception (Shannon *et al.*, 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 the number of acoustic channels (Shannon *et al.*, 1995). In quiet, normal hearing (NH) subjects reach asymptote with about eight channels (Dorman *et al.*, 1997), although more channels are needed when listening in noise (Dorman *et al.*, 1998). NH subjects listening to six-channel simulations perform similarly to CI users (Dorman and Loizou, 1998). Although limited spectral information is sufficient for high levels of consonant, vowel, and sentence perception, acoustic signals containing complex spectra, such as music, may require well over 30 channels to be perceived accurately (Shannon *et al.*, 2004; Shannon, 2005).

[a] Author to whom correspondence should be addressed; electronic mail: jeremyloebach@gmail.com
[b] Electronic mail: tbent@indiana.edu
[c] Electronic mail: pisoni@indiana.edu

Compared to perception of linguistic information in the speech signal, considerably less is known about the perception of indexical information in both CI users and NH subjects listening to CI simulations. Cleary and Pisoni (2002) demonstrated that prelingually deafened pediatric CI users have more difficulty discriminating talkers based on their voices than do NH children. Deaf children with cochlear implants, who could discriminate between talkers, performed comparably to NH children, but all CI users required much larger pitch differences between talkers than NH controls in order to successfully distinguish talkers (see also Cleary *et al.*, 2005). Moreover, the ability to discriminate talkers was found to be significantly correlated with several different speech perception outcome measures, indicating that the acoustic information necessary for talker discrimination is closely coupled with the information used in speech perception and spoken word recognition (Cleary *et al.*, 2005).

Compared to talker discrimination, the identification of a talker's gender is easier for both NH subjects listening to eight-channel CI simulations and CI users (Fu *et al.*, 2004). Both groups of listeners may be relying on the pitch of the fundamental frequency to classify speaker gender (Fu *et al.*, 2005). For NH subjects, the performance on gender identification tasks depends 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 and requires fewer spectral channels for sinewave than noise vocoders (Gonzales and Oliver, 2005). Given that sinewave vocoders may encode temporal information with higher fidelity (Gonzales and Oliver, 2005), it is possible that subjects are utilizing temporal cues to discriminate the fundamental frequency.

The explicit identification of speakers by their voice alone may require more spectral detail compared to talker or gender discrimination, and subjects may be using different perceptual processes for linguistic as compared to indexical tasks (Vongphoe and Zeng, 2005). When presented with vowels produced by ten different talkers, CI users and NH subjects listening to eight-channel simulations performed equally well on vowel identification but CI users performed significantly more poorly when identifying the speakers by voice alone (Vongphoe and Zeng, 2005). When considered on a talker-by-talker basis, however, the primary source of errors in talker identification was between talkers with higher pitch voices (adult females, girls, and boys) (Vongphoe and Zeng, 2005). When boys and girls were excluded, the CI users resembled the normal hearing subjects listening to eight-channel simulations, just as they did in the vowel identification task. This alternative interpretation suggests that both linguistic and indexical tasks may rely on similar perceptual processing strategies, but when the ranges of vocal pitch overlap substantially, CI users may not be receiving sufficient spectral information to reliably identify talkers. This possibility, however, has not been experimentally addressed.

### B. Perceptual learning in linguistic and indexical tasks

Understanding the perceptual interactions between indexical and linguistic properties of speech is important for a number of reasons. A talker's specific realizations of acoustic-phonetic parameters will ultimately determine their speech intelligibility (Bond and Moore, 1994; Bradlow *et al.*, 1996; Cox *et al.*, 1987; Hood and Poole, 1980), and adult listeners are constantly adjusting their internal categories to accommodate new talkers (e.g., Eisner and McQueen, 2005). Such perceptual learning, which can be defined as long-term changes in the perceptual system based on sensory experiences that influences future behaviors and responses (Goldstone, 1998; Fahle and Poggio, 2002), may play a central role in adaptation to novel talkers and speaking styles. Moreover, familiarity with a talker's voice has been shown to provide a significant benefit when listening to speech in noise (Nygaard *et al.*, 1994; Nygaard and Pisoni, 1998), indicating interactions between the linguistic and indexical channels of information, and suggesting that they may indeed be encoded in the same stream.

Perceptual learning of speech can be both talker independent and talker dependent. When a listener is explicitly trained to classify an ambiguous sound in a real word, category boundaries for other words containing the sound will be adjusted to accommodate the new pronunciation by that particular talker (Eisner and McQueen, 2005). Moreover, if the manipulated stimuli have a greater degree of potential generalizabillity (e.g., a voicing distinction for alveolars that could apply to bilabials and velars as well), perceptual learning will be robust and occur independent of talker (Kraljic and Samuel, 2006). Talker-independent perceptual learning has also been shown to be beneficial for listeners with extensive experience listening to synthetic speech produced by rule (Schwab *et al.*, 1985; Greenspan *et al.*, 1988), time-compressed speech (Dupoux and Green, 1997), foreign accented speech (Bradlow and Bent, 2008; Clarke and Garrett, 2004; Weil, 2001), and noise-vocoded speech (Davis *et al.*, 2005). Critically, the benefits of perceptual learning extend to new talkers and new speech signals created using the same types of signal degradation.

Training can have large effects on the perceptual learning of CI simulated speech. The types of tasks and stimulus materials used during training have been shown to affect post-training gains and generalization. Robust generalization and transfer of perceptual learning to novel meaningful sentences has been demonstrated for individual phonemes (Fu *et al.*, 2005), words (Loebach and Pisoni, 2008), meaningful and anomalous sentences (Burkholder, 2005; Davis *et al.*, 2005; Loebach and Pisoni, 2008), and environmental sounds (Loebach and Pisoni, 2008). Although previous research has examined the perception of indexical information in CI simulated speech, it is unknown whether training on talker identification will generalize to word or sentence recognition under conditions of severe spectral degradation, as has been shown for naturally produced speech (Nygaard and Pisoni, 1998).

### C. The present study

Understanding whether indexical and linguistic tasks differentially affect perceptual learning will help to elucidate whether these sources of information are encoded in the same stream and utilize similar perceptual mechanisms for processing. Moreover, understanding how linguistic and indexical channels of information interact in speech perception may provide novel insights into possible training methodologies for new CI users. The present study compared how training on linguistic and indexical tasks affected listeners' ability to accurately perceive words in sentences. Using sentences processed with an eight-channel sinewave vocoder, NH subjects were trained to identify either the gender or identity of six talkers or transcribe their speech. Pre- to post-training comparisons of transcription accuracy scores assessed the effectiveness of training. Given the results of the previous studies, we hypothesized that subjects trained on talker identification would perform better than those who were trained on gender identification due to increased attentional demands required during training. Moreover, we predicted that training on talker identification would match or exceed the performance of subjects trained using a conventional sentence transcription task because of the controlled attention required to learn to identify the talkers from severely degraded stimuli.

## II. METHODS

### A. Subjects

Seventy-eight young adults participated in the study (60 female, 18 male; mean age, 21 years). All subjects were native speakers of American English and reported having normal hearing at the time of testing. Subjects were recruited from the Indiana University community and received monetary compensation ($10/session) or course credit (1 credit/session) for their participation. Of the 78 subjects tested, 6 were excluded from the final data analysis (2 failed to return for the generalization session, 1 failed to return in a timely manner, and 3 due to program malfunctions).

### B. Stimuli

Stimuli consisted of 124 meaningful [76 high predictability (HP) and 48 low predictability (LP)] and 48 anomalous (AS) speech in noise test (SPIN) sentences (Kalikow *et al.* 1977; Clopper and Pisoni, 2006). SPIN sentences are balanced for phoneme occurrence in English and contain between five and eight words, the last of which is the keyword to be identified. The final word in the HP sentences is 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 retained the 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 for the familiarization portion of the experiment. Wav files of the materials were obtained from the Nationwide Speech Corpus (Clopper and Pisoni, 2006). Materials were produced by six speakers (three male, three female) from the midland dialect.

### C. Synthesis

All stimuli were processed using Tiger CIS to simulate an eight-channel cochlear implant using the CIS processing strategy. Stimulus processing involved two phases: an analysis phase, which used band pass filters to divide the signal into eight nonlinearly spaced channels (between 200 and 7000 Hz, 24 dB/octave slope) and a low pass filter to derive the amplitude envelope from each channel (400 Hz, 24 dB/octave slope); and a synthesis phase, which replaced the frequency content of each channel with a sinusoid that was modulated with its matched amplitude envelope. All training, familiarization, and testing materials were processed with the eight-channel vocoder unless otherwise noted: a subset 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 16-bit Windows PCM wav files (22 kHz sampling rate) and normalized to 65 dB rms (LEVEL v2.0.3) (Tice and Carrell, 1998) to ensure that all stimuli were of equal intensity and to eliminate any peak clipping.

### D. 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 computers that were equipped with 15-in.-color liquid crystal display (LCD) monitors. Audio signals were presented over Beyer Dynamic DT-100 headphones and were calibrated with a voltmeter to a 1000 Hz tone at 70 dB v 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. Each trial was preceded by a 1000 ms silent interval, followed by a fixation cross presented at the center of the screen for 500 ms to alert the subject to the upcoming trial. The subject was prompted to record their response after stimulus offset. For the transcription trials, a dialog box was presented on the screen signaling 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. Each experimental session lasted approximately 40–60 min.

#### *1. Training*

Training took place over two sessions. The tasks and materials varied across blocks but the same block structure was used for all groups. All stimuli were randomized within each block. Session 1 began with a pretest to establish a

TABLE I. Tasks and stimulus materials in the pretest, familiarization, and training blocks of session 1. Feedback was only provided in the training blocks

| Pretest | Familiarization | Training (feedback) |
|---|---|---|
| Transcribe:<br>30 LP sentences<br>30 AS sentences | Passively listen:<br>Rainbow passage | Transcribe<br>ID talker<br>ID gender:<br>150 HP sentences |

baseline level of performance before training (see Table I) and expose subjects to the processing condition in order to provide an unbiased assessment of training efficacy. In the pretest blocks, subjects transcribed 30 unique spectrally degraded LP sentences, followed by 30 unique spectrally degraded AS sentences. In these blocks, the subjects simply transcribed the sentences and received no feedback.

During the familiarization phase, subjects passively listened to spectrally degraded versions of the Rainbow passage produced by each of the six talkers in order to become familiar with the voices and synthesis condition and teach them the appropriate labels that would be used 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, or Sue). Subjects were told that they would be asked to identify the talkers by name and they should 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 these were paired with the appropriate gender label (male or female) for each talker. These subjects were told that they would be asked to identify the gender of the talkers, and they should listen carefully for any information that would help them learn to recognize each talker's gender. Finally, subjects in the transcription group heard each passage presented along with the name of the talker who produced it. However, they were told that they would be asked to transcribe sentences produced by each talker and they should listen carefully in order to better understand the degraded signals.

The three training blocks in session 1 consisted of 150 spectrally degraded HP sentences (the same 25 sentences produced by each talker). During training, subjects were presented with a sentence and asked to make a response appropriate for their training group. All subjects were provided with feedback regardless of the accuracy of their responses. Subjects in the talker ID group identified the talker by clicking one of six buttons on the computer screen labeled with the talkers' names. After the subject recorded his/her 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 recorded his/her 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. They received the correct transcription of the sentence as feedback.

Session 2 (Table II) was completed within 3 days of session 1 and began with a repetition of the familiarization phase in which subjects again heard the Rainbow Passage produced by each talker. The purpose of this block was to refamiliarize the listener with the voices and labels, since at least 24 h had passed since the first training session. Two training blocks followed, consisting of 90 HP sentences (15 sentences produced by each talker). Again, subjects received feedback regardless of their response.

Transfer of training to novel materials that were more spectrally degraded than the training materials was then assessed, and subjects were asked to transcribe 36 new HP sentences that were more severely spectrally degraded than the stimuli that the subjects heard during training (18 four-channels, 18 six channels). Generalization of training was tested in the following two blocks, and subjects were asked to transcribe 18 novel LP sentences and 18 novel AS sentences. Following generalization, two posttest blocks assessed the relative gains in performance due to training. Subjects transcribed 12 AS sentences and 12 LP sentences selected from the pretest blocks. We chose to conduct the posttest last in order to more accurately assess the benefits from training. In particular, we wanted to rule out the influence of procedural learning, which could distort the posttest performance (since subjects in the talker or gender ID groups had less experience with the sentence transcription task).

### 2. Analysis and scoring

Keyword accuracy scores were based on the final word in each sentence. Common misspellings and homophones were counted as correct responses. However, 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 posttest comparisons provided an assessment of the gains from training across the three training groups. Comparison of performance at pre- and posttest to performance on new materials provided an assessment of generalization of training to novel

TABLE II. Tasks and stimulus materials in the refamiliarization, training, generalization, and posttest blocks of session 2. Feedback was only provided in the training blocks.

| Familiarization | Training (feedback) | Generalization: Novel materials | Transfer: More degraded | Posttest |
|---|---|---|---|---|
| Passively listen:<br>Rainbow passage | Transcribe<br>ID talker<br>ID gender:<br>90 HP sentences | Transcribe:<br>18 LP sentences<br>18 AS sentence | Transcribe:<br>18 HP (four-band SVS)<br>18 HP (six-band SVS)<br>sentences | Transcribe:<br>12 AS sentences<br>12 LP sentences<br>(from pretest) |

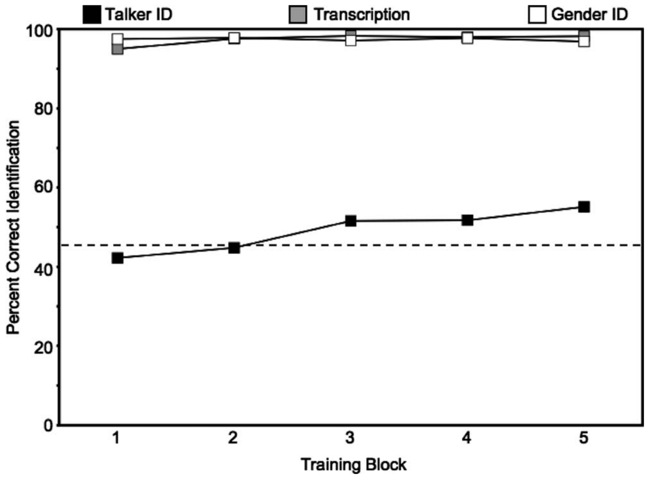

FIG. 1. Perceptual learning across the five training blocks. The dashed horizontal line indicates the level of performance that subjects must exceed in order to be considered significantly different from chance in the talker identification condition. Subjects trained to transcribe the sentences (Transcription)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

stimuli. Generalization was said to have occurred if performance was significantly higher than the pretest scores and greater than or equal to posttest scores. Comparisons of the performance on the four- and six-channel stimuli provided an assessment of how well training transferred to more severely degraded stimuli.

## III. RESULTS

### A. Perceptual learning during training

Accuracy on the training tasks varied by training group (Fig. 1). Subjects in the gender ID and transcription training groups performed near ceiling, whereas subjects in the talker ID group performed just above chance.

Subjects in the transcription training group performed extremely well across all five training blocks, starting at 95% in block 1 and ending at 98% in block 5. A univariate analysis of variance (ANOVA) revealed a significant main effect of block $[F(4,190)=6.441,\ p<0.001]$, indicating that subjects showed an improvement across training blocks. *Post hoc* Bonferroni tests revealed that subject performance in block 1 was significantly lower than performance in all other blocks (all $p<0.009$), which 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. The 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.

The performance of the talker ID group was considerably more variable across subjects. Since intergender confusions (identifying male talkers as female, or female talkers as male) were rare (<2%), a more conservative level of chance was used (33.3% rather than 16.7%). According to the binomial probability distribution, performance must exceed 44.5% correct to be significantly greater than chance. Most subjects ($n=26$) were able to identify talkers at a level greater than chance beginning in block 2 and showed continued improvement as training progressed (block 1: 42.2%, block 2: 44.8%, block 3: 51.6%, block 4: 51.7%, and block 5: 55.1%). A subset of talkers ($n=5$) could never identify talkers at a level greater than chance and their data were analyzed separately. 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 $[F(1,250)=39.509,\ p<0.001]$. Subjects identified female talkers (54%) more accurately than male talkers (44%).

### B. Performance after training

#### *1. Pretest, posttest, and generalization*

An *omnibus* repeated measures analysis of variance was conducted on the data across the three experimental phases (pretest, posttest and generalization) using training condition (transcription, talker ID, and gender ID) as between subjects variables and sentence type (meaningful versus anomalous) and speaker gender (male versus female) as within subjects factors. A highly significant main effect of training was observed $[F(2,252)=10.918,\ p<0.001]$, indicating that subjects' performance was influenced by the type of training that they experienced. *Post hoc* Bonferroni tests revealed that subjects trained in the talker ID task performed as well as subjects trained in the transcription task ($p=1.00$) and both groups performed significantly better than subjects in the gender ID training task (both $p\leqslant 0.001$). The effect of sentence type was not significant $[F(1,252)=2.962,\ p=0.086]$. Performance did not differ between anomalous and meaningful sentences. A significant main effect of talker gender was observed $[F(1,252)=53.276,\ p<0.001]$. Subjects performed better on sentences produced by female talkers than on sentences produced by male talkers. None of the two-way and three-way interactions reached statistical significance (all $p>0.05$).

Using the findings from the *omnibus* ANOVA as motivation, individual univariate ANOVAs were then conducted on the data in each individual experimental phase to explore these effects in more detail (Fig. 2). For the pretest data, a significant main effect of training was observed $[F(2,252)=4.38,\ p=0.013]$, indicating differences across the three groups before training began. *Post hoc* Bonferroni tests revealed that subjects in the transcription group ($M=55.7\%$, SD=12.3) performed as well as subjects in the talker ID group ($M=51.9$, SD=14.7, $p=0.069$) and significantly better than subjects in the gender ID group ($M=50.7$, SD=13.2, $p=0.016$). However, subjects in the talker ID and gender ID groups did not differ from one another at pretest ($p=1.00$). A significant main effect of sentence type was also observed $[F(1,252)=45.880,\ p<0.001]$. On average, subjects performed better on the anomalous sentences ($M=57.4$, SD

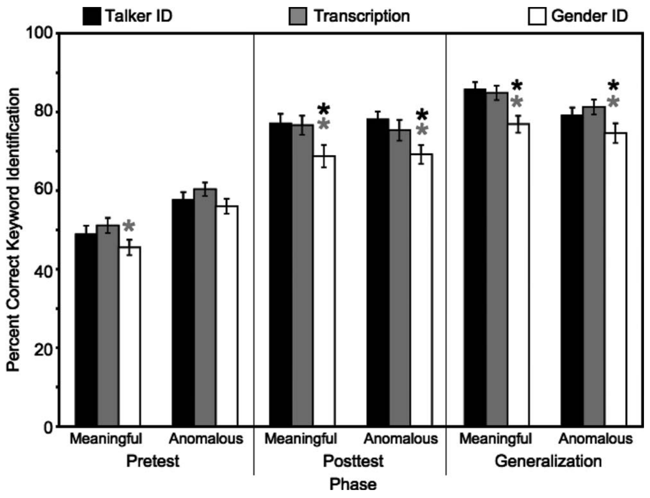

FIG. 2. Percent correct keyword identification scores for subjects trained on talker identification (talker ID), gender identification (gender ID), and sentence transcription (Transcription) on the pre-, post-, and generalization tests. Asterisks appearing over a bar indicate significant differences in performance between that group and another group (indicated by color).

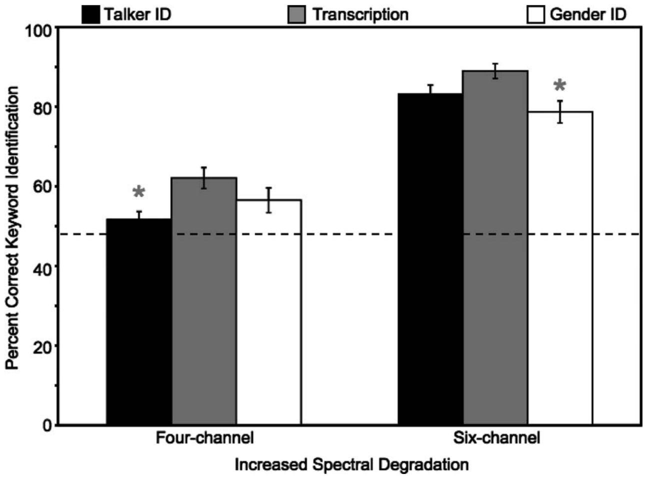

FIG. 3. Percent correct keyword identification scores for subjects trained on talker identification (talker ID), gender identification (gender ID), and sentence transcription (transcription) on the more severe spectral degradation condition. Colored asterisks appearing over a bar indicate significant differences in performance between that group and another group.

=12.0) than meaningful sentences ($M$=48, SD=13.6). Presumably, this effect is driven by the fact that all subjects received meaningful sentences first and that the better performance on the anomalous sentences (9%) is likely attributable to rapid adaptation to the processing conditions (Davis *et al.*, 2005). A significant main effect of talker gender was also observed [$F(1,252)=59.06$, $p<0.001$]. Subjects performed better on materials produced by female talkers ($M$ =58.2, SD=12.7) than by male talkers ($M$=47.2, SD=12.3). None of the interactions were statistically significant, except for the two-way interaction between sentence type and talker gender [$F(1,252)=10.437$, $p<0.001$]. This effect may reflect the general performance differences on meaningful and anomalous sentences, since subjects always performed better on female talkers as compared to male talkers.

Given the main effect of training on the pretest data, comparisons of performance across groups on the posttest data must control for pretest scores (to ensure that the performance at posttest is not merely a factor of better performance at pretest). A univariate ANOVA was conducted on the posttest data specifying pretest scores as a covariate. A significant main effect was observed for pretest scores [$F(1,251)=28.441$, $p<0.001$], confirming that pretest scores differed across groups and indicating that the performance at pretest influenced performance at posttest. Despite the effect of the covariate, a significant main effect of training was still observed for the posttest scores [$F(2,251)$ $=5.305$, $p=0.005$], indicating that training differentially affected performance at posttest (Fig. 2). *Post hoc* Bonferroni tests revealed that subjects in the transcription group ($M$ =76, SD=16.9) performed as well as subjects in the talker ID group ($M$=76.4, SD=15.5, $p$=1.00), and that both groups performed significantly better than subjects in the gender ID group ($M$=68.9, SD=16.7, both $p<0.01$). The main effect of sentence type was not statistically significant [$F(1,252)$ $=0.130$, $p=0.719$]. On average, subjects performed equally well on anomalous sentences ($M$=74.5, SD=15.3) and meaningful sentences ($M$=73.6, SD=17.3). A significant main effect of talker gender was observed [$F(1,252)$ $=15.715$, $p<0.001$]. Subjects performed better on materials produced by female talkers ($M$=78.2, SD=15.6) as compared to male talkers ($M$=69.9, SD=16.1). None of the interactions were statistically significant, except for the two-way interaction between sentence training and talker gender [$F(2,252)=4.548$, $p=0.011$]. This effect may reflect the general performance differences provided by training, since across all three training groups, subjects always performed better on female talkers as compared to male talkers.

A univariate ANOVA on the generalization data also revealed a significant main effect of Training [$F(2,252)$ $=8.53$, $p<0.001$], indicating that training differentially affected performance during generalization to novel materials (Fig. 2). *Post hoc* Bonferroni tests revealed that subjects in the talker ID group ($M$=82.4, SD=13.6) performed as well as subjects in the transcription group ($M$=82.9, SD=12, $p$ =1.00), and both groups performed significantly better than subjects in the gender ID group ($M$=75.6, SD=14.7, both $p<0.001$). A significant main effect of sentence type was also observed [$F(1,252)=6.718$, $p=0.01$], indicating that subjects performed better on the meaningful sentences ($M$ =82.6, SD=12.3) than anomalous sentences ($M$=78.3, SD =14.1). A significant main effect of talker gender was also observed [$F(1,252)=19.96$, $p<0.001$], indicating that subjects performed better on female talkers ($M$=84, SD=13.5) than male talkers ($M$=76.9, SD=13.4). None of the interactions were significant.

### *2. Transfer of training to increased spectral degradation*

Subjects showed a graded response to stimuli that were more severely spectrally degraded (Fig. 3). Overall, subjects were more accurate at transcribing sentences in the six-channel processing condition (transcription: 83.1%; gender ID: 78.6%; talker ID: 88.9%) than sentences in the four-channel processing condition (transcription: 51.7%; gender

ID: 56.4%; talker ID: 61.9%). Comparison of the performance on the four-channel processed sentences across the 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 the performance on the six-band stimuli across the 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.

### C. Talker ID training subgroups

An additional finding of the present study emerged when we first assessed the 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%). However, a small subset of subjects were unable to identify talkers at a level greater than chance. 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 training blocks (1=30.8%, 2=35.4%, 3=36.3%, 4=34.7%, and 5=32.9%), as indicated by a univariate ANOVA [$F(4,40)=0.05$, $p=0.628$]. 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 able to learn the talker identification task. A series of one-way ANOVAs revealed that performance did not differ at pretest for either 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 pretest, their performance increased at a different rate depending on how well they performed in the training task. These findings are not likely to be caused by inattention or laziness 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 (see Cleary and Pisoni, 2002; Cleary *et al.*, 2005).

## IV. DISCUSSION

The present study assessed whether training tasks that have different attentional requirements produce equivalent levels of perceptual learning, generalization, and transfer to new materials and tasks. Although all three types of training produced significant pre- to posttest gains in performance, talker ID and sentence transcription training appeared to provide the largest and most robust improvements (Fig. 2). Generalization to new stimulus materials was equivalent for the talker ID and transcription training groups, both of whom performed significantly better than the subjects trained on gender ID (Fig. 2). Generalization to materials that were more spectrally degraded showed a mixed pattern of results (Fig. 3). For stimuli that were more spectrally degraded (four and six channel), subjects trained on sentence transcription performed best, subjects trained on gender ID performed worst, and subjects trained on talker ID displayed an intermediate level of performance.

Two main conclusions can be drawn from these data. First, training on explicit indexical tasks can yield equivalent levels of perceptual learning and transfer compared to training using traditional transcription tasks if task demands are high enough to require sustained attention and deeper processing. Evidence for this conclusion comes from the across training group comparisons of posttest and generalization scores for subjects in the talker ID group who performed similarly to the subjects in the transcription training group and significantly better than the subjects in the gender ID training group (Fig. 2). Compared to gender ID training (in which subjects were at ceiling in the first training block), talker ID training is a more difficult task under CI simulations, requiring high levels of controlled attention and deeper processing. The gains observed from training on indexical tasks also suggest that when a listener is exposed to a speech signal that is meaningful in their native language they cannot help interpreting it in a linguistically significant manner. Although subjects' controlled attention in the talker and gender ID tasks was not directed toward the linguistic information in the signal, they still processed the linguistic content of the sentences automatically [similar to the effects seen in the well known Stroop effect (Stroop, 1935)]. The degree to which they did so appears to be mediated by the specific training task they were asked to carry out.

Second, the benefits of training may be determined by whether the subject can successfully access the acoustic information in the speech signal and the depth of perceptual processing required to succeed in the training task. Subjects in the talker ID group, who had to make fine acoustic-phonetic distinctions among voices and hence process the signal more deeply, performed significantly better than subjects in the gender ID group. Moreover, the poor performing 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. Taken together, these findings suggest that the access and attention to fine acoustic-phonetic details learned during talker ID training may enhance a listener's ability to extract linguistic information from the speech signal (see also, Nygaard *et al.*, 1994).

### A. Transfer of indexical training to linguistic tasks: Interactions of transfer appropriate processing and levels of processing

The findings of the present study suggest that perceptual learning of spectrally degraded speech can be facilitated by processes used to carry out indexical tasks even though the specific tasks that subjects perform at training and testing are fundamentally different. The transfer appropriate processing (TAP) theory of learning and memory predicts that performance will be maximized when the task used during testing is the same as the task used during training (e.g., Morris *et al.*, 1977; Roediger *et al.*, 1989). Under the TAP framework, it would be expected that subjects in the transcription training group would receive the largest benefit from training, since the task they carried out during training (sentence transcription) was the same task that they were asked to carry out at posttest and generalization. This expectation was only partially supported. Although subjects in the transcription group performed best overall, their performance was equivalent to the subjects trained on talker ID (except for the four-channel generalization test), suggesting that a factor other than TAP influenced performance, particularly for subjects in the talker ID group.

The levels of processing (LoP) approach to learning and memory suggests that tasks that require deeper analysis and processing will yield better long term recall (Craik and Lockhart, 1972). Talker identification under a CI simulation is considerably more difficult than for natural speech. The acoustic information that specifies the voice of the talker in the unprocessed signal appears to be significantly degraded when processed through a CI speech processor, requiring more controlled attention and deeper processing. Gender identification is much easier, suggesting that the acoustic information needed to successfully identify the gender of a talker is relatively well preserved (e.g., Gonzales and Oliver, 2005). Therefore, the task demands placed on a listener are significantly higher in a talker ID task than those in a gender ID task (which requires only shallow processing). Thus, under the LoP framework, subjects in the talker ID training condition should be expected to perform better than the subjects in the gender ID group, since the latter requires considerably more detailed acoustic analysis and hence deeper processing. This expectation was supported.

The effects of TAP and LoP during training are particularly relevant to our understanding of perceptual learning in speech. The data from the present study suggest that explicit indexical training tasks can produce robust transfer to linguistic tasks despite the predictions under the TAP framework. If the training task is difficult enough to require sustained controlled attention and deep processing, transfer will be equivalent to that produced by conventional linguistic training tasks. The differences in performance between the talker ID and gender ID training conditions support this hypothesis. We chose to use a longer training period (2 days and over 240 sentences) in order to get a more accurate and stable estimation of perceptual learning due to training. Subjects in the gender ID training group performed near ceiling from the first training block, suggesting that their task was easier and required shallower processing. Subjects in the talker ID training group, however, performed more poorly, only improving above chance at the end of the second training block, and still showing evidence of improvement in the fifth and final training blocks. The task demands placed on these subjects required deeper processing, further indicating that differences in controlled attention across training conditions differentially affects perceptual learning.

These data also suggest that training tasks in which subjects have room for improvement should produce better (and more robust) perceptual learning since the subjects are constantly being challenged to improve their performance (Bjork, 1994). Subjects in the transcription and gender ID training groups performed at ceiling in the first block of training and did not have the opportunity to improve their performance. This ceiling effect may have been a result of the ease of the binary decision for the subjects in the gender ID task, or by the fact that each of the high predictability sentences was repeated six times in each block (once by each talker), creating a familiarity effect for subjects in the transcription training group. Subjects in the talker ID training group were given a much more difficult task and showed significant improvement across training blocks. Thus, the differences in performance across the training groups at posttest and generalization may be influenced not only by the demands of the training task but also by the potential room for improvement on the tasks. Greater gains in performance may have been observed for subjects in the transcription and gender ID training groups if the tasks afforded greater opportunity for improvement. Both of these tasks could be made more difficult by utilizing male and female talkers that have closer fundamental frequencies, making the gender identification task more difficult, and by using lower predictability sentences that are not repeated, making the transcription task more difficult. Overall, optimal training tasks should require more controlled attention and deeper processing but also allow substantial room for improvement.

One possible concern with the present research is the evaluation of the gains from training, and whether these are true training effects or simple exposure effects (since subjects in the gender ID and transcription were at ceiling during training). Although subjects in all training groups showed a significant pre-to posttest improvement, the amount of improvement that one would expect from merely being exposed to the materials in the absence of feedback is unknown particularly for indexical tasks. If these were simple exposure effects, however, we would expect posttest scores to be equal across all training groups (since all subjects were exposed to the same materials). This was not the case as subjects in the talker ID and transcription training groups improved their performance significantly more than subjects in the gender ID training group, indicating that these differences in performance cannot be attributed to simple exposure effects. Moreover, the gender ID task provided an internal control condition since discrimination of speaker gender is an easy task using vocoded speech, requiring shallower processing and less controlled attention. Since gains obtained from training were significantly higher for subjects in the transcription and talker ID groups than subjects in the gender ID group, we can infer differential enhancement of perceptual learning by explicit indexical training tasks and do not need to consider a separate control group who was merely exposed to the materials. Additionally, previous research suggests that subjects improve by roughly 10% over time for sentence transcription tasks without any feedback (Davis *et al.*, 2005). Subjects in the present study exceeded this figure, improving on average 20%, indicating that pre-/posttest improvement was well above what would be expected from mere exposure alone.

## B. Access to the acoustic information in the signal

An additional finding in this study was the correlation between the ability to learn to identify talkers by voice and performance in the sentence transcription tasks. Although the vast majority of subjects (84%) could learn to identify the talkers at a level greater than chance, those who could not performed significantly worse in the posttest, generalization, and transfer blocks. When considered together, the findings from the present study 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. Understanding why some listeners perform poorly on talker identification despite training will require further study.

Previous research has shown that subjects, who can successfully learn to explicitly identify novel talkers by voice, display higher word identification accuracy scores in noise when compared to subjects, who could not learn to identify talkers by voice (Nygaard *et al.*, 1994). The findings of the present study replicate these earlier findings using spectrally degraded vocoded speech, suggesting that the ability to successfully encode and retain talker-specific acoustic information in memory affects perceptual learning of degraded speech. Other research has shown that pediatric CI users who could accurately discriminate talkers by voice had higher word identification scores as compared to children who could not (Cleary *et al.*, 2005). The findings of the present study replicate these findings in normal hearing subjects listening to CI simulations. When considered together, these data provide additional converging evidence for the interaction of lexical and indexical information in speech perception, and suggest that the two streams may indeed be encoded and processed together (Pisoni, 1997).

Although indexical information was traditionally thought to be encoded separately from linguistic information (see Abercrombie, 1967; Halle, 1985), the two streams of information interact at a fundamental perceptual level. Speech perception has been viewed as a talker independent process, where the listener must "normalize" the acoustic information across talkers in order to extract the context-free symbolic linguistic content from the signal (see Pisoni, 1997 for a review). While listeners do adjust internal linguistic categories to accommodate new talkers (Eisner and McQueen, 2005; Kraljic and Samuel, 2006), such indexical information does not appear to be lost or discarded following linguistic interpretation (Nygaard *et al.*, 1994, Nygaard and Pisoni, 1998). The present set of results suggest that the earliest stages of speech perception may be episodic and highly context-dependent in nature, with the listener encoding detailed indexical information along with the linguistic information and retaining both types of information well after the original sensory trace has decayed (Pisoni, 1997; Goldinger, 1998). These findings suggest that indexical and linguistic information are encoded in the same stream and interact bidirectionally to influence perception. The degree to which such indexical information is utilized, however, appears to depend on the task, listener, and specific stimulus materials.

## C. Behavioral and clinical implications

The findings from the present study suggest the existence of multiple routes to the perceptual learning of speech. Although previous training studies utilize traditional methods of training that exclusively focus the listener's attention on the abstract symbolic linguistic content encoded in the speech signal (e.g., Fu *et al.*, 2005b), other routes to perceptual leaning can yield equivalent outcomes and benefits. The crucial factor seems to be the amount of controlled attention that is required of the subject and the depth of perceptual processing required to succeed in the training task. Indexical processing tasks that require significant amounts of controlled attention and deep processing (e.g., talker identification) can be just as effective as tasks that rely exclusively on explicit attention to the linguistic content of the message. This finding has important implications for training and rehabilitation paradigms for hearing impaired listeners who receive cochlear implants and hearing aids. The benefit obtained in the present study suggests that a variety of tasks and stimulus materials could be utilized effectively to maximize perceptual learning after cochlear implantation, thereby increasing outcome and benefit. Explicit training and instruction on how to distinguish and identify individual voices may provide the CI user with a more stable foundation for voice recognition that can generalize to new talkers in novel listening environments (such as voice tracking in noise). Additionally, including a wide variety of stimulus materials and challenging perceptual tasks may promote interest in training and reduce boredom and fatigue.

Although the overall goal of cochlear implantation has been to restore receptive auditory capacity for speech, there are many other nonlinguistic aspects to hearing on which a CI user could experience benefit. Sound localization, the detection and identification of ecologically significant environmental sounds, and the enjoyment of music are all aspects of normal hearing that have not been fully explored in CI users. Because all of these tasks require attention to nonlinguistic acoustic information, a greater variety in training tasks and materials may yield robust outcomes across multiple domains, many of which may also produce additional gains in speech perception and spoken language processing. If the goal of cochlear implantation is to provide the user with access to the acoustic world, perceptual learning and training paradigms for cochlear implant users should not be limited exclusively to conventional linguistic tasks that rely on word recognition and linguistic interpretation of the speech signal.

## ACKNOWLEDGMENTS

This work was supported by NIH-NIDCD Training Grant No. T32-DC00012 and NIH-NIDCD Research Grant No. R01-DC00011. The authors would like to thank Althea Bauernschmidt and Larry Phillips for their assistance in data collection.

Abercrombie, D. (**1967**). *Elements of General Phonetics* (Aldine, Chicago).

Bond, Z. S., and Moore, T. J. (**1994**). "A note on the acoustic-phonetic characteristics of inadvertently clear speech," Speech Commun. **14**, pp. 325–337.

Bradlow, A. R., and Bent, T. (**2008**). "Perceptual adaptation to non-native speech," Cognition **107**, 707–729.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**, 255–272.

Bjork, R. A. (**1994**). "Memory and metamemory considerations in the training of human beings," in *Metacognition: Knowing about Knowing*, edited by J. Metcalfe and A. Shimamura (MIT, Cambridge), pp. 185–205.

Burkholder, R. A. (**2005**). "Perceptual learning of speech processed through an acoustic simulation of a cochlear implant," Ph.D. thesis Indiana University.

Clarke, C. M., and Garrett, M. F. (**2004**). "Rapid adaptation to foreign-accented English," J. Acoust. Soc. Am. **116**, 3647–3658.

Cleary, M., and Pisoni, D. B. (**2002**). "Talker discrimination by prelingually deaf children with cochlear implants: Preliminary results," Ann. Otol. Rhinol. Laryngol. Suppl. **111**, 113–118.

Cleary, M., Pisoni, D. B., and Kirk, K. I. (**2005**). "Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants," J. Speech Lang. Hear. Res. **48**, 204–223.

Clopper, C. G., and Pisoni, D. B. (**2006**). "The Nationwide Speech Project: A new corpus of American English dialects," Speech Commun. **48**, 633–644.

Craik, F. I. M., and Lockhart, R. S. (**1972**). "Levels of processing: A framework for memory research," J. Verbal Learn. Verbal Behav. **11**, 671–684.

Cox, R. M., Alexander, G. C., and Gilmore, C. (**1987**). "Development of the connected speech test (CST)," Ear Hear. **8**, 119–126.

Davis, M. H., Johnsrude, 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," J. Exp. Psychol. Gen. **134**, 222–241.

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 Hear. **19**, 162–166.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Simulating the effect of cochlear-implant electrode insertion depth on speech understanding," J. Acoust. Soc. Am. **102**, 2993–2996.

Dorman, M., Loizou, P., Fitzke, J., and Tu, Z. (**1998**). "The recognition of sentences in noise by normal hearing listeners using simulations of cochlear implant signal processors with 6-20 channels," J. Acoust. Soc. Am. **104**, 3583–3585.

Dupoux, E., and Green, K. P. (**1997**). "Perceptual adjustment to highly compressed speech: Effects of talker and rate changes," J. Exp. Psychol. Hum. Percept. Perform. **23**, 914–927.

Eisner, F., and McQueen, J. M. (**2005**). "The specificity of perceptual learning in speech processing," Percept. Psychophys. **67**, 224–238.

Fairbanks, G. (**1940**). *Voice and Articulation Drillbook* (Harper and Row, New York).

Fahle, M., and Poggio, T. (**2002**). *Perceptual Learning* (MIT Press, Cambridge).

Fu, Q-J., Chinchilla, S., and Galvin, J. J. (**2004**). "The role of spectral and temporal cues on voice gender discrimination by normal-hearing listeners and cochlear implant users," J. Assoc. Res. Otolaryngol. **5**, 253–260.

Fu, Q-J., Chinchilla, S., Nogaki, G., and Galvin, J. J. (**2005a**). "Voice gender identification by cochlear implant users: The role of spectral and temporal resolution," J. Acoust. Soc. Am. **118**, 1711–1718.

Fu, Q-J., Galvin, J. J., Wang, X., and Nogaki, G. (**2005b**). "Moderate auditory training can improve speech performance of adult cochlear implant patients," ARLO **6**, 106–111.

Goldinger, S. D. (**1998**). "Echoes of echoes? An episodic theory of lexical access," Psychol. Rev. **105**, 251–279.

Goldstone, R. L. (**1998**). "Perceptual learning," Annu. Rev. Psychol. **49**, 585–612.

Gonzales, J., and Oliver, J. C. (**2005**). "Gender and speaker identification as a function of the number of channels in spectrally reduced speech," J. Acoust. Soc. Am. **118**, 461–470.

Greenspan, S. L., Nusbaum, H. C., and Pisoni, D. B. (**1988**). "Perceptual learning of synthetic speech produced by rule," J. Exp. Psychol. Learn. Mem. Cogn. **14**, 421–433.

Halle, M. (**1985**). "Speculation about the representation of words in memory," in *Phonetic Linguistics*, edited by V. Fromkin (Academic, New York), pp. 101–114.

Hood, J. D., and Poole, J. P. (**1980**). "Influence of the speaker and other factors affecting speech intelligibility," Audiology **19**, 434–55.

Kalikow, D. N., Stevens, K. N., and Elliot, L. L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.

Kraljic, T., and Samuel, A. S. (**2006**). "Generalization in perceptual learning for speech," Psychon. Bull. Rev. **13**, 262–268.

Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**(1), 98–104.

Loebach, J. L., and Pisoni, D. B. (**2008**). "Perceptual learning of spectrally degraded speech and environmental sounds," J. Acoust. Soc. Am. **123**, 1126–1139.

Morris, C. D., Bransford, J. D., and Franks, J. J. (**1977**). "Levels of processing versus transfer appropriate processing," J. Verbal Learn. Verbal Behav. **16**, 519–533.

Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (**1994**). "Speech perception as a talker-contingent process," Psychol. Sci. **5**, 42–46.

Nygaard, L. C., and Pisoni, D. B. (**1998**). "Talker-specific learning in speech perception," Percept. Psychophys. **60**, 355–376.

Pisoni, D. B. (**1997**). "Some thoughts on 'normalization' in speech perception," in *Talker variability in speech processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego), 9–32.

Roediger, H. L., Weldon, M. S., and Challis, B. H. (**1989**). "Explaining dissociations between implicit and explicit measures of retention: A processing account," in *Varieties of Memory and Consciousness: Essays in honor of Endel Tulving*, edited by H. L. Roediger and F. I. M. Craik (Hillsdale, Erlbaum), pp. 3–41.

Schwab, E. C., Nusbaum, H. C., and Pisoni, D. B. (**1985**). "Some effects of training on the perception of synthetic speech," Hum. Factors **27**, 395–408.

Shannon, R. V. (**2005**). "Speech and music have different requirements for spectral resolution," Int. Rev. Neurobiol. **70**, 121–134.

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 Oto-Laryngol., Suppl. **552**, 1–5.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Stroop, J. R. (**1935**). "Studies of interference in serial verbal reactions," J. Exp. Psychol. **18**, 643–662.

Tice, R., and Carrell, T. (**1998**). LEVEL 16 V2.0.3, University of Nebraska, Lincoln, NE.

Vongphoe, M., and Zeng, F. G. (**2005**). "Speaker recognition with temporal cues in acoustic and electric hearing," J. Acoust. Soc. Am. **118**, 1055–1061.

Weil, S. A. (**2001**). "Foreign accented speech: Adaptation and generalization," Master's thesis, Ohio State University.

http://www.tigerspeech.com