

# Perceptual adaptation and intelligibility of multiple talkers for two types of degraded speech

Tessa Bent<sup>a)</sup>

*Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue, Bloomington, Indiana 47405*

Adam Buchwald

*Department of Speech-Language Pathology and Audiology, New York University, 665 Broadway, Suite 910, New York, New York 10012*

David B. Pisoni

*Department of Psychological and Brain Sciences, Indiana University, 1101 East 10th Street, Bloomington, Indiana 47405*

(Received 7 July 2008; revised 29 July 2009; accepted 31 July 2009)

Talker intelligibility and perceptual adaptation under cochlear implant (CI)-simulation and speech in multi-talker babble were compared. The stimuli consisted of 100 sentences produced by 20 native English talkers. The sentences were processed to simulate listening with an eight-channel CI or were mixed with multi-talker babble. Stimuli were presented to 400 listeners in a sentence transcription task (200 listeners in each condition). Perceptual adaptation was measured for each talker by comparing intelligibility in the first 20 sentences of the experiment to intelligibility in the last 20 sentences. Perceptual adaptation patterns were also compared across the two degradation conditions by comparing performance in blocks of ten sentences. The most intelligible talkers under CI-simulation also tended to be the most intelligible talkers in multi-talker babble. Furthermore, listeners demonstrated a greater degree of perceptual adaptation in the CI-simulation condition compared to the multi-talker babble condition although the extent of adaptation varied widely across talkers. Listeners reached asymptote later in the experiment in the CI-simulation condition compared with the multi-talker babble condition. Overall, these two forms of degradation did not differ in their effect on talker intelligibility, although they did result in differences in the amount and time-course of perceptual adaptation. © 2009 Acoustical Society of America.

[DOI: 10.1121/1.3212930]

PACS number(s): 43.71.Gv [RSN]

Pages: 2660–2669

## I. INTRODUCTION

Although it is well known that talkers differ in intelligibility (Bond and Moore, 1994; Bradlow *et al.*, 1996; Hazan and Markham, 2004; Hood and Poole, 1980), less is known about the stability of these differences across different types of signal degradation. In this paper, we ask whether the talkers who are highly intelligible in multi-talker babble are also highly intelligible when their speech is processed by a cochlear implant (CI) simulator. In addition, we investigate whether some talkers are easier or harder to adapt to. Finally, we report on how the type of degradation contributes to the process of adaptation.

### A. Speech intelligibility

What factors determine speech intelligibility?<sup>1</sup> Broadly speaking, traditional views of speech intelligibility have maintained that intelligibility is a property of the speaker, the acoustic signal, or of the specific words being perceived (Black, 1957; Bond and Moore, 1994; Bradlow *et al.*, 1996;

Hood and Poole, 1980; Howes, 1952, 1957). There is empirical evidence supporting each of these views. For example, certain properties of words (e.g., segmental composition, length, and frequency) have been shown to influence speech intelligibility (Black, 1957; Howes, 1952, 1957). Similarly, it has been shown that several specific acoustic properties of a talker's speech (e.g., speaking rate and vowel dispersion) play a crucial role in determining speech intelligibility (Bond and Moore, 1994; Bradlow *et al.*, 1996; Hood and Poole, 1980), as can properties of the listening environment (e.g., Assman and Summerfield, 2004; Fletcher and Steinberg, 1924; Miller, 1947; Miller and Nicely, 1955). However, the speech materials and the talker are not the only relevant factors in determining speech intelligibility. Instead, a variety of research findings suggest that speech intelligibility is also influenced by properties of the listener (e.g., Bent and Bradlow, 2003; Imai *et al.*, 2003; Labov and Ash, 1997; Mason, 1946) and linguistic context (e.g., Healy and Montgomery, 2007), as well as interactions among these factors (e.g., Moore, 2003; Rogers *et al.*, 2006).

Whether differences across talkers are maintained under different listening environments is an issue that has not been extensively studied. In one of the few extant studies, Cox *et al.* (1987) found that relative intelligibility rankings

---

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: tbent@indiana.edu

among six talkers were generally maintained across four levels of noise degradation (speech mixed with babble). More recently, Green *et al.*, (2007) reported no differences in talker intelligibility among three groups of listeners: normal-hearing listeners, CI listeners, and simulated CI listeners. The stimuli included single words appended to a carrier phrase from six talkers and semantically-anomalous sentences from two talkers. The stimuli were produced by two adult male, two adult female, and two child female talkers. Each of these three groups contained one high intelligibility talker and one low intelligibility talker (based on earlier results from Hazan and Markham, 2004). These stimulus materials were presented to CI users and normal-hearing listeners. Normal-hearing listeners heard the speech either mixed with multi-talker babble at a favorable signal-to-noise ratio or under CI simulation. Green *et al.* (2007) reported that intelligibility was relatively consistent across listeners and degradation types. These two studies suggest that at least some talker characteristics that promote intelligibility are beneficial across listener populations and listening conditions. The present study continues this line of inquiry by comparing the intelligibility of speech mixed with multi-talker babble to the intelligibility of CI-simulated speech for a larger number of talkers.

## B. Perceptual learning

Several recent studies have found that listeners demonstrate both talker-dependent and talker-independent perceptual learning of speech (e.g., Bradlow and Bent, 2008; Norris *et al.*, 2003). With respect to talker-dependent learning, as listeners become more familiar with a talker's voice their word recognition accuracy increases (Bradlow and Bent, 2008; Nygaard *et al.*, 1994; Nygaard and Pisoni, 1998). These studies use overall intelligibility to assess adaptation, so it was not possible to determine how the specific experience with the talkers' voices enabled the listeners to improve their ability to identify those talkers' words. Other work using synthetic manipulations for specific phoneme contrasts suggests that listeners adjust their phonemic category boundaries in talker-specific ways (e.g., Eisner and McQueen, 2005; Norris *et al.*, 2003). These studies suggest that using lexical knowledge, listeners shift their category boundaries as needed for particular talkers.

The effect of linguistic experience has also been found to be talker-independent. Talker-independent learning has been shown for adjustments to phoneme category boundaries for native-accented speech (Kraljic and Samuel, 2006, 2007). Furthermore, a beneficial effect of experience on speech intelligibility has been shown for listeners with extensive experience listening to foreign accented speech (Bradlow and Bent, 2008; Clarke and Garrett, 2004; Weil, 2001), speech produced by talkers with hearing impairments (McGarr, 1983), speech synthesized by rule (Schwab *et al.*, 1985; Greenspan *et al.*, 1988), and computer manipulated speech (Dupoux and Green, 1997; Pallier *et al.*, 1998). Critically, this benefit has been reported to extend to new talkers and to new speech signals created using the same types of signal degradation (Bradlow and Bent, 2008; Francis *et al.*, 2007;

Greenspan *et al.*, 1988; McGarr, 1983). Of particular relevance to the current study, listeners also show rapid adaptation to noise-vocoded speech (Davis *et al.*, 2005; Hervais-Adelman *et al.*, 2008). In the study of Davis *et al.* (2005) perceptual adaptation occurred without feedback across 30 sentences and was enhanced by feedback either in the form of an orthographic presentation of the stimulus or repetition of an unprocessed version of the stimulus. Adaptation was stronger with meaningful sentences than non-word sentences although adaptation with words was the same with real words and non-words (Hervais-Adelman *et al.*, 2008). Other training studies have also shown that the amount of perceptual learning depends on the type of materials listeners are trained with (Loebach and Pisoni, 2008).

Investigating listeners' adaptation to new talkers and the conditions that allow for this adaptation provides valuable information about the robustness and extent of plasticity in the speech perception system. Furthermore, uncovering the conditions that are most beneficial to adaptation can potentially help in the development of training programs for listeners with speech perception difficulties such as listeners with hearing impairment or second language learners. We addressed this type of perceptual adaptation in the present study by comparing performance on an initial group of sentences in a novel listening condition to performance after the listener has been exposed to the condition for many sentences. The results of previous studies suggest that we should observe significantly better performance after exposure to a novel listening condition. Furthermore, we investigated performance over the time-course of the experiment to determine the asymptote for adaptation by calculating performance in blocks of ten sentences (i.e., performance in the first ten sentences, second ten sentences, etc.).

While the studies reviewed above on perceptual learning of speech have demonstrated a great deal of flexibility of listeners' perceptual systems, they have typically focused on only one talker and only one type of signal degradation. In the current study, we addressed these gaps in two ways. First, we investigated adaptation across a large number of talkers to determine the extent of variation in perceptual adaptation to different talkers. Second, we compared how these differences in perceptual adaptation to different talkers may be affected by two different types of signal degradation. One type of signal degradation, CI-simulated speech, was selected because the perceptual adaptation of CI users is a topic that is still relatively unexplored, and attempts to understand their perceptual adaptation should be helpful in creating training protocol for individuals with CIs; synthesizing speech with a CI simulator for unimpaired subjects is a useful tool for addressing the issues with this population (Dorman and Loizou, 1998; Dorman *et al.*, 1997; Shannon *et al.*, 1995). The second type of signal degradation, mixing speech with multi-talker babble, was selected to be an ecologically valid degradation method that would provide a comparison with the CI-simulated speech; this will allow us to address whether individuals adapt to all types of signal degradation in the same way by determining whether the speakers to whom adaptation is more robust are the same in each condition. These forms of degradation are similar in that they both

make spectral detail less accessible. However, vocoding and the addition of multi-talker babble degrade the signal in different ways: spectral broadening and masking, respectively.

### C. The present study

In this paper, we report on an investigation of how talker characteristics interact with degradation type to determine speech intelligibility and perceptual adaptation. One of the aims of this experiment was to determine whether and how inter-talker differences in intelligibility change depending on the type of degradation (i.e., CI-simulated speech versus speech mixed with multi-talker babble). The second aim of this study was to investigate how across talker differences and signal degradation type affect perceptual adaptation. Understanding how the interaction of talker characteristics and listening environment influences intelligibility and perceptual adaptation is an important goal in identifying the factors that contribute to speech perception and learning.

In the current experiment, intelligibility scores for ten male and ten female talkers were compared under two listening conditions: CI-simulation and multi-talker babble. Listeners were presented with speech from only one talker in one listening condition. Four hundred listeners were tested in total: 200 listeners for each listening condition. Intelligibility scores were compared across listening conditions, and the extent of adaptation to the speech over the time-course of the experiment was assessed.

## II. METHOD

### A. Stimuli

The experimental materials were sentences taken from the Indiana Multi-talker Sentence Database (Karl and Pisoni, 1994). This database includes recordings of 100 Harvard sentences (IEEE, 1969) produced by 20 talkers (10 male and 10 female), with a total of 2000 sentences. All talkers were speakers of general American English. The sentences were processed in two ways to assess speech intelligibility under CI-simulated listening conditions and when mixed with multi-talker babble.

### B. CI-simulation

For the CI-simulation condition, each sentence was processed through an eight-channel sinewave vocoder using the CI simulator TIGERCIS (<http://www.tigerspeech.com/>). 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. The eight-channel simulation was chosen because on average normal-hearing listeners perform similar to CI users when listening to eight-channel simulations compared to greater or fewer numbers of channels (Dorman *et al.*, 1997). Furthermore, a sine-wave vocoder was employed rather than noise-band vocoder be-

cause sine-wave vocoders also approximate CI user performance more closely than noise-band vocoders (Gonzalez and Oliver, 2005). However, it should be noted that this simulation is not an entirely accurate representation of the information presented to CI users. Specifically, due to the spectral side-bands around the sine-wave carriers, more information regarding the fundamental frequency is available in the simulation than is through a CI. The availability of this information may affect the intelligibility of speech in ways that are not representative of CI processing.

### C. Multi-talker babble

For the multi-talker babble condition, the original sentences were mixed with six-talker babble (three male and three female talkers) at a signal-to-noise ratio of 0. The same babble file was used for each of the 2000 sentences. None of the talkers included in the babble file were the same as the target talkers. This signal-to-noise ratio was chosen based on pilot data in which the intelligibility of the sentences mixed with multi-talker babble was matched with intelligibility of the eight-channel CI-simulated sentences. The speech in this condition was not processed.

### D. Participants

Four hundred normal-hearing listeners participated in this study (268 females and 132 males with an average age of 21.4 years). All listeners were native speakers of English and reported no current speech or hearing impairments at the time of testing. The majority of the participants were from the mid-west and indicated their place of birth as Indiana ( $n=191$ ), Illinois ( $n=50$ ), Ohio ( $n=13$ ), Michigan ( $n=13$ ), Minnesota ( $n=6$ ), Missouri ( $n=6$ ), Wisconsin ( $n=3$ ), Iowa ( $n=3$ ), or Kansas ( $n=2$ ). The remaining participants were from the south ( $n=40$ ), northeast ( $n=29$ ), west ( $n=21$ ), the U.S., state not specified ( $n=6$ ), or outside of the U.S. ( $n=7$ ). Ten participants did not indicate their place of birth. Most of the participants did not speak a foreign language, but 22 indicated knowing one language other than English. These languages included Spanish ( $n=8$ ), Urdu ( $n=3$ ), Chinese ( $n=1$ ), French ( $n=1$ ), German ( $n=1$ ), Italian ( $n=1$ ), Korean ( $n=1$ ), Polish ( $n=1$ ), Japanese ( $n=1$ ), Swedish ( $n=1$ ), and Arabic ( $n=1$ ). Two of the participants indicated knowing two foreign languages: Hebrew/Spanish ( $n=1$ ) and Bengali/Hindi ( $n=1$ ). Listeners were either paid \$5.00 for their participation or received course credit in an introductory psychology course. Participants were undergraduate students at Indiana University or members of the greater Bloomington community. In the CI-simulation condition, four subjects' data were removed because they were determined to be outliers (their keyword correct score was at least three standard deviations below the mean for that talker). Their data were replaced by data from four new listeners.

### E. Experimental task

In each condition, a talker's intelligibility was assessed by examining the performance of ten normal-hearing listeners on a sentence transcription task (20 talkers  $\times$  2 degradation conditions  $\times$  10 listeners = 400 listeners total). Each lis-

tener was presented with speech from one condition (i.e., CI-simulation or multi-talker babble) and heard only one talker during the course of the experiment, allowing us to assess differences in adaptation across talkers. During testing, each participant wore Beyer Dynamic DT-100 headphones while sitting in front of a Power Mac G4. Each sentence was played over the headphones followed by a dialog box presented on the screen, which prompted the listener to type what he or she heard. Each sentence was presented once in a randomized order. The experiment was self-paced so participants could take as long as needed to enter a response. Listeners were not provided with feedback as to the accuracy of their responses. Prior to the first experimental trial, participants were familiarized with the type of degradation by hearing two familiar nursery rhymes (“Jack and Jill” and “Star Light, Star Bright”) produced by a talker not included in the Hoosier Multi-Talker Sentence Database or in the multi-talker babble, which had been processed in the same manner as the sentences in their experimental condition. During familiarization, listeners were not required to make any responses.

### F. Scoring

The responses were scored based on number of keywords correct. Each test sentence has five keywords. Keywords were only counted as correct if all and only the correct morphemes were present. Words with added or deleted morphemes were counted as incorrect. Obvious misspellings and homophones were counted as correct.

## III. RESULTS

The results under the two types of degradation were compared in several ways. First, intelligibility across talkers under the two types of degradation was compared in order to determine whether high and low intelligibility talkers in one condition are also the high and low intelligibility talkers in the other condition. Second, male speakers were directly compared with the female speakers in terms of intelligibility; gender was shown to be a significant predictor of intelligibility under quiet listening conditions (Bradlow *et al.*, 1996). Third, we examined the extent of perceptual adaptation under each experimental condition by comparing performance on the first 20 sentences with performance on the last 20 sentences. Differences in perceptual adaptation between the two conditions were also compared. We compared performance in ten blocks of ten sentences each to assess performance over the time-course of the experiment to investigate the rate of perceptual adaptation.

### A. Comparison of intelligibility between the CI-simulation and multi-talker babble conditions

The intelligibility scores from the two conditions, CI-simulation and multi-talker babble, were compared. The keyword accuracy scores for the CI-simulated condition and the multi-talker babble condition were significantly correlated ( $r=0.73$ ,  $p<0.001$ ). Talkers who were highly intelligible under one type of degradation, CI-simulation, also tended to also be highly intelligible under the other type of degrada-



FIG. 1. Comparison of keyword intelligibility for the two degradation conditions, CI-simulated speech and speech mixed with multi-talker babble. Intelligibility scores under these two conditions were significantly correlated.

tion, multi-talker babble. A scatterplot of the keyword intelligibility scores in the two degradation conditions is shown in Fig. 1. It should be noted that different listeners were used for the two conditions. Therefore, potentially confounding listener variables were introduced (i.e., differences in dialect and other linguistic experiences across listeners).

The intelligibility scores for each talker in the CI-simulation condition and the multi-talker babble conditions were also compared to intelligibility scores in the quiet (gathered by Karl and Pisoni, 1994). Here *sentence* intelligibility was considered rather than *keyword* intelligibility as Karl and Pisoni (1994) only reported sentence intelligibility scores (due to a lack of variation in keyword correct scores). Intelligibility scores in quiet were not significantly correlated with intelligibility in the CI-simulation condition ( $r=0.35$ ,  $ns$ ) and were not significantly correlated with intelligibility in multi-talker babble condition ( $r=0.36$ ,  $ns$ ). However, it should be noted that the range of intelligibility scores in the quiet was relatively small.

### B. Gender differences

The data from both the CI-simulation and multi-talker babble conditions revealed that female talkers were more intelligible than male talkers. In the CI-simulation condition, female talkers (mean=84%, SD=11) were significantly more intelligible than male talkers [mean=77%, SD=11;  $t(198)=4.61$ ,  $p<0.001$ ]. Similarly, female talkers (mean=81%, SD=14) were more intelligible than male talkers [mean=65%, SD=13;  $t(198)=8.47$ ,  $p<0.001$ ] in the multi-talker babble condition. The gender difference in quiet, shown previously in Bradlow *et al.* (1996) with the same talkers, is maintained under the two forms of signal degradation tested here.

### C. Perceptual adaptation

In addition to overall speech intelligibility, adaptation to the speech in each condition was assessed by examining improvement from the first 20 sentences to the last 20 sentences, a measure of perceptual adaptation. For the CI-simulation condition, this analysis revealed significant adaptation, with significantly more keywords correct in the

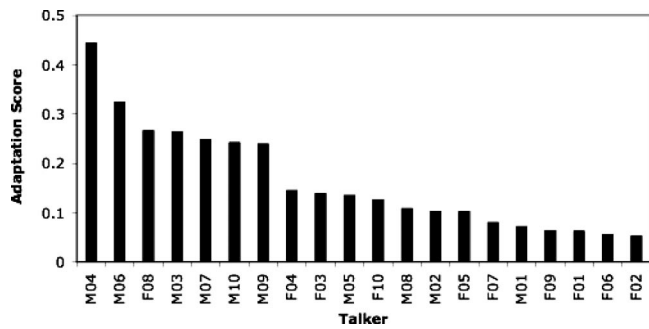


FIG. 2. Adaptation scores for the CI-simulated listening conditions. Talkers are ordered on the x-axis from left to right by their adaptation scores in the CI-simulation condition. While listeners adapted to the speech from all talkers, the extent of adaptation depended on the particular talker.

last 20 sentences (mean=84%, SD=11) than in the first 20 sentences [mean=73%, SD=15;  $t(19)=16.6$ ,  $p < 0.001$ ]. Thus, listeners adapted to the CI-simulated speech without explicit feedback. An adaptation score was also calculated by subtracting the keywords correct in the first 20 sentences from the keywords correct in the last 20 sentences divided by keywords correct in the first 20 sentences. While listeners adapted to all talkers, a great deal of variation was observed in the extent of adaptation across individual talkers, with adaptation scores ranging from 0.05 to 0.44 for individual talkers. These data are shown in Fig. 2.

As with the CI-simulation condition, the perceptual adaptation analysis with the data from the multi-talker babble condition also revealed rapid adaptation, with significantly more keywords correct in the last 20 sentences (mean =75%, SD=13) than in the first 20 sentences [mean=69%, SD=16;  $t(19)=6.45$ ,  $p < 0.001$ ]. Again, listeners rapidly adapted to the speech without explicit feedback. A great deal of variation was also observed in the extent of adaptation for the talkers, with adaptation scores ranging from 0.00 to 0.30 for individual talkers. These data are shown in Fig. 3.

In addition to assessing perceptual adaptation in each condition, we also compared the extent of adaptation in the two degradation conditions using the adaptation scores. A paired  $t$ -test revealed that listeners showed greater perceptual adaptation in the CI-simulated listening condition (mean =0.16, SD=0.11) than in the multi-talker babble condition

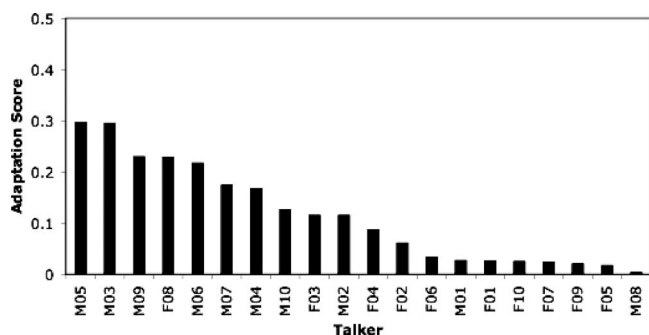


FIG. 3. Adaptation scores for speech mixed with multi-talker babble. The talkers are ordered on the x-axis based on their adaptation scores in the multi-talker babble condition, decreasing from left to right. While listeners adapted to the speech from all talkers, the extent of adaptation depended on the particular talker.

[mean=0.12, SD=0.10;  $t(19)=2.68$ ,  $p=0.015$ ]. Furthermore, when comparing each talker's adaptation scores between the two degradation conditions, the extent of adaptation in the two conditions was significantly correlated ( $r=0.68$ ,  $p = 0.001$ ).

The vast majority of individual listeners showed improvement over the time-course of the experiment. In the CI-simulation condition, 91% of listeners showed adaptation scores above zero while in the multi-talker babble condition slightly fewer individuals showed improvement with 84% of adaptation scores above zero. Of the listeners who did not show adaptation across the course of the experiment (9% of listeners in the CI-simulation condition and 16% of listeners in the multi-talker babble condition), some listeners were at near ceiling level within the first 20 sentences, leaving little room for improvement over the time-course of the experiment. In the CI-simulation condition 7 of the 18 listeners who did not show perceptual adaptation were at 90% correct or above in the first 20 sentences and 12 of the 32 listeners showing no improvement in the multi-talker babble condition were at 90% or better in the first 20 sentences.

In addition to investigating performance during the beginning and the end of the experiment, we also examined the pattern of perceptual adaptation across the entire experiment by comparing performance for each block of ten sentences (i.e., Block 1=first ten sentences, Block 2=second ten sentences, etc.). This analysis allowed us to determine the point at which listeners reached asymptote in each condition and to compare the shape of the adaptation curves in each condition. The perceptual adaptation curves are shown in Fig. 4.

A repeated-measures analysis of variance (ANOVA) was conducted on these data with block as the within-subject repeated measure and condition (CI-simulation versus multi-talker babble) as the between-subjects variable. Results revealed main effects of block [ $F(9,398)=77.84$ ,  $p < 0.001$ ] and condition [ $F(1,398)=29.61$ ,  $p < 0.001$ ] as well as an interaction between block and condition [ $F(9,398)=4.46$ ,  $p < 0.05$ ]. Because we found a significant interaction, separate repeated measures ANOVAs were conducted on the data from the two conditions.

For the CI-simulation condition, the effect of block was highly significant [ $F(9,199)=60.96$ ,  $p < 0.001$ ]. *Post-hoc* pairwise comparisons with Bonferroni correction were made between each of the blocks in order to determine the asymptote. From these comparisons, listeners reached asymptote at the sixth block of sentences. That is, performance in the sixth block of sentences was not significantly different from any later blocks in the experiment, which also did not differ from one another. While listeners showed considerable adaptation across the first 60 sentences in the experiment (starting at 70% correct in the first ten sentences with gains to 83% correct in the sixth block of sentences), there was no further improvement observed after the sixth block (performance in the tenth block was only 1% higher than in the sixth block).

For the multi-talker babble condition, the effect of block was also highly significant [ $F(9,199)=23.18$ ,  $p < 0.001$ ]. Pairwise comparisons revealed that listeners reached asymptote earlier in the experiment in this condition compared to the CI-simulation condition. In the multi-talker babble con-

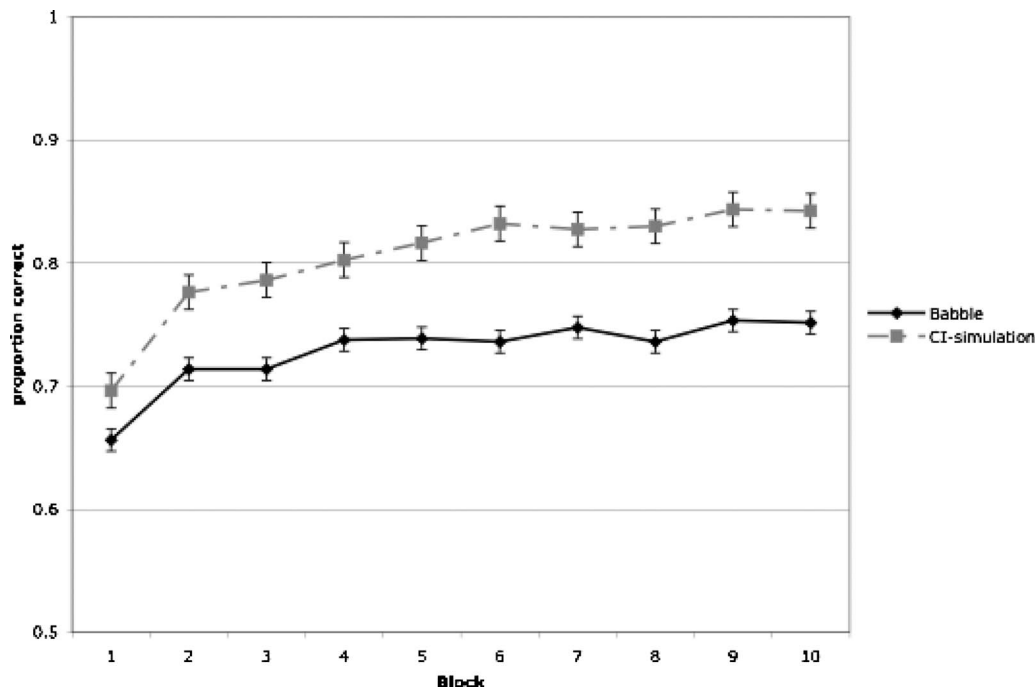


FIG. 4. Perceptual adaptation curves for the CI-simulation condition and multi-talker babble condition. On the x-axis performance for the ten blocks of sentences is shown (each composed of ten sentences). The y-axis displays proportion correct for keywords.

dition, Block 4 did not differ from any of the blocks later in the experiment, which also did not differ from one another. Listeners showed considerable adaptation in the first 40 sentences in the experiment (starting at 67% correct in the first ten sentences with gains to 74% in the fourth block of sentences). However, listeners showed little additional adaptation later in the experiment as performance only increased 1% from Block 4 (74%) to Block 10 (75%).

Comparisons were also made between the two conditions for each block using independent samples *t*-tests. Because of the large number of *t*-tests, Bonferroni correction was applied which indicated that *p*-values must be 0.005 or less to be considered significant. Comparisons across the two conditions showed that performance in the first block of ten sentences was not different in the two conditions [ $t(398) = 2.12$ , *ns*]. However, performance was significantly higher in the CI-simulation condition compared to the multi-talker babble condition in Blocks 2–10 ( $p \leq 0.001$ ).

While the above analyses collapsed across male and female talkers, we wanted to investigate how learning across the experiment was affected by talker gender. Figure 5 shows learning across the time-course of the experiment in the two conditions divided by gender. It becomes clear that both initial and final performances are least accurate for male talkers when heard in multi-talker babble.

Results from the analyses of the perceptual adaptation revealed mostly similarities and some differences across the two degradation conditions. Adaptation scores across talkers were correlated for the two degradation conditions, and the adaptation effect was very robust. On average, listeners showed adaptation for nearly all talkers. Moreover, nearly all listeners showed adaptation over the time-course of the experiment. While the adaptation scores were correlated between the two conditions, listeners showed greater adapta-

tion in the CI-simulation condition than in the multi-talker babble condition and showed the least accurate performance across the experiment for the male talkers in the multi-talker babble condition. Moreover, listeners reached asymptote later in the CI-simulation condition compared with the multi-talker babble condition.

#### IV. GENERAL DISCUSSION

Results from the current study suggest that across-talker differences in speech intelligibility are maintained across two types of signal degradation. Talkers who were found to be highly intelligible under CI-simulation were also highly intelligible when their speech was presented in multi-talker babble. Our findings support the recent conclusions of [Green et al. \(2007\)](#) who suggest that inter-talker differences are maintained across different listener groups (i.e., CI users, normal-hearing listeners presented with speech in a low level of babble or with CI-simulated speech). The present results replicated their earlier findings in a larger talker sample using sentence length materials. The overall patterns in our study diverge from previous studies examining relative intelligibility among talkers from different language backgrounds, indicating that speech intelligibility rankings may change depending on listener language background ([Bent and Bradlow, 2003](#); [Imai et al., 2003](#); [van Wijngaarden, 2001](#); [van Wijngaarden et al., 2002](#); cf. [Major et al., 2002](#); [Munro et al., 2006](#)). However, we suspect that some factors such as language background may result in stronger talker-listener interactions compared with other factors such as hearing loss. If this is the case, then native listeners from the same speech community—regardless of their hearing status—will find the same talkers most intelligible, but listeners from different language backgrounds, especially na-

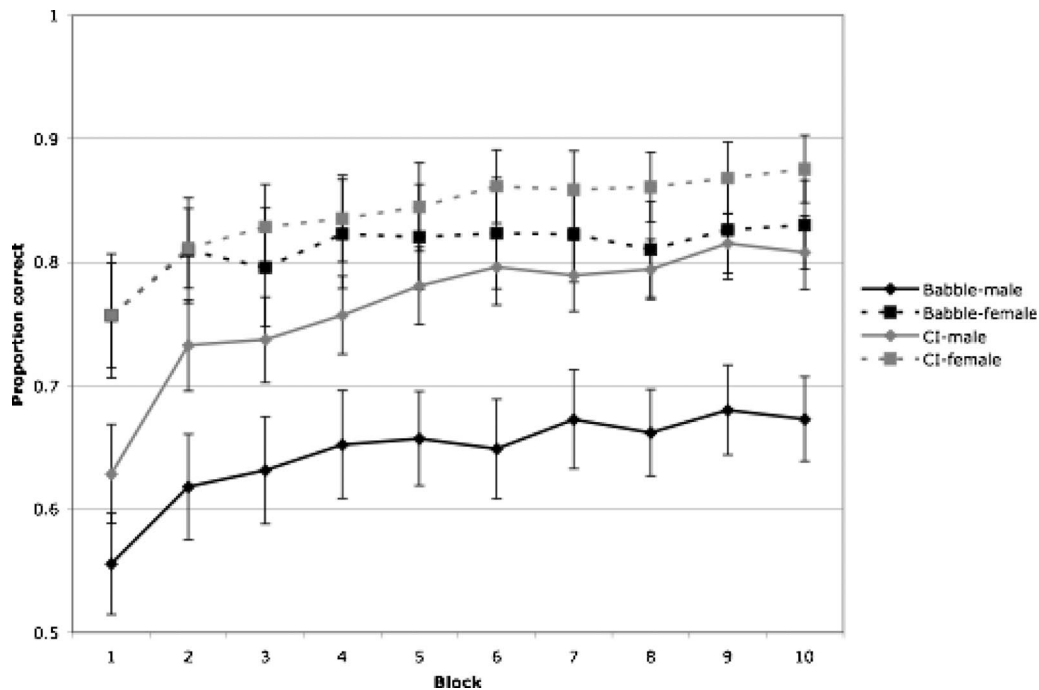


FIG. 5. Perceptual adaptation curves for the CI-simulation condition and multi-talker babble condition divided into male and female talkers. On the x-axis performance for the ten blocks of sentences is shown (each composed of ten sentences). The y-axis displays proportion correct for keywords.

tive and non-native listeners, may find different talkers most intelligible. The results from the present study reveal that intelligibility under multi-talker babble listening conditions is correlated with intelligibility under CI-simulation. However, as both of the degradation types tested here make the spectral detail in the speech signal less available, the extent to which this result can be generalized to other types of degradation remains an empirical issue. It should also be noted that different listeners were included in the two degradation conditions. As previous studies have found that factors about the listener can influence intelligibility, it may be the case that the results found here were influenced by across-listeners differences. A match in linguistic experiences of the talker and listener could have enhanced intelligibility scores for certain talkers, whereas a mismatch could have caused an intelligibility decrement. However, all listeners were native speakers of American English, and all talkers were speakers of general American English. Recent findings on the intelligibility of different American English dialects suggest that there is not an interaction between the dialect of the talker and the listener (Clopper and Bradlow, 2008).

Although mixing speech with multi-talker babble is typically considered an ecologically valid process for degrading speech, it should be noted that the same recordings—collected in quiet conditions—were used in the quiet and multi-talker babble listening conditions. Therefore, modifications that talkers make when they speak in noisy environments (e.g., Lombard speech: Junqua, 1993; Lane and Tranel, 1971; Lane et al., 1970; Lombard, 1911; Summers et al., 1988) were not performed in these recordings. In general, when listening in noise, speech produced in noise tends to be more intelligible than speech produced in the quiet (Summers et al., 1988). Furthermore, certain talkers are more effective at making modifications and adjustments that

help listeners in noisy environments when they are producing speech with noise present. Moreover, females generally tend to produce more intelligible Lombard speech than males (Junqua, 1993). Similarly, some talkers are better at making their speech highly intelligible when asked to speak clearly for listeners with hearing loss compared to the intelligibility of their speech when asked to speak conversationally (Ferguson, 2004; Picheny et al., 1985; Uchanski et al., 1996).

The remainder of this section explores two issues raised by the data reported here. In particular, we address the issues of perceptual adaptation and the observed gender differences.

### A. Perceptual adaptation

Listeners with normal hearing are able to quickly adapt and accurately perceive speech under a variety of different listening conditions. In the present experiment, the analysis of adaptation to the degraded speech revealed the flexibility of the speech perception system. Even in the absence of any feedback, listeners recognized the talker's utterances more accurately after several minutes of exposure to the experimental stimuli (i.e., last 20 sentences) compared to the beginning of exposure to these stimuli (i.e., first 20 sentences). The extent of perceptual adaptation varied for each talker and in each type of signal degradation. It should be noted that for both degradation conditions, each listener in the experiment was only exposed to the speech of one talker. Therefore, the adaptation observed in the experiment is likely a result of adaptation to talker specific characteristics as well as adaptation to the degradation condition.

Listeners showed greater perceptual adaptation in the CI-simulation condition than in the multi-talker babble condition. However, the correlation between the talkers' adaptation scores in the two degradation conditions was positive

and significant. One likely source of the greater adaptation in the CI-simulation condition compared to the multi-talker babble condition is the novelty of the former type of degradation. The listeners in the current study had never experienced CI-simulated listening conditions before participating in the experiment, whereas listeners have had experience perceiving speech in environments with competing talkers. Thus, listeners are already practiced at picking out a given talker in noisy listening environments that are similar to the multi-talker babble condition, and must only adapt to the specifics of the multi-talker babble added to the speech in the experiment. Listeners are unlikely to learn a new listening strategy in this experiment, whereas the exposure to CI-simulated speech provided in this experiment was their first experience with this form of degradation. Therefore, they may have been able to acquire a new listening strategy during the course of the exposure to the CI-simulated speech. Evidence for this hypothesis comes from the finding that listeners continued to learn further into the experiment (i.e., they reached asymptote in the sixth block) in the CI-simulation condition compared with the multi-talker babble condition (i.e., they reached asymptote in the fourth block of sentences). The initial steep gains seen in both conditions may be a result of procedural learning while later learning may be a consequence of perceptual learning involving learning to better extract information from the degraded stimuli (Francis and Nusbaum, 2002). Another reason for the greater adaptation in the CI-simulation condition compared with the multi-talker babble condition is that the manipulation in the CI-simulation condition is a less variable and more predictable form of degradation than the multi-talker babble condition. Once a listener learns how the speech had been degraded in the CI-simulation condition, she can reliably use this information to more successfully interpret future utterances. In contrast, the way the multi-talker babble interacts with the target speech stimulus changes from sentence to sentence, which may hinder a listener's ability to apply knowledge learned from one sentence to the next. However, since the babble file that was mixed with the speech was the same from trial to trial, listeners may have been able to generalize their knowledge of the specifics of the babble noise from sentence to sentence (see Felty *et al.*, 2009). Although listeners showed robust learning for talkers in both conditions, the performance at the beginning and end of the experiment for the male talkers in the multi-talker babble condition was significantly lower than for female talkers in multi-talker babble or talkers from either gender in the CI-simulation condition.

It is worth noting that the listeners in this experiment did not receive any feedback, which suggests that they may have taken advantage of semantic and syntactic cues to enable them to learn how to perceive the speech under the two degradation conditions. Results from Davis *et al.* (2005) demonstrate that greater learning is observed in cases with meaningful sentences compared with non-word sentences, in which all words are non-words, or Jabberwocky sentences, in which only content words are replaced with non-words but real English function words remain. Davis *et al.* (2005) also tested perceptual learning of CI-simulated speech without

feedback but only assessed 30 sentences. The present results add to their earlier findings by demonstrating that listeners continue to learn up through exposure to 60 sentences with eight-channel vocoded speech but then reach asymptote and show no further learning on the final 40 sentences. For listeners to achieve further gains, the inclusion of appropriate feedback would presumably be necessary.

In terms of generalizing about perceptual adaptation differences between babble and CI-simulation, it is worth noting that the experimental design employed here was a between-subjects design in which participants were exposed to only one type of signal degradation. While this methodological choice allowed us to fully explore the time-course and extent of perceptual adaptation to a given talker with a specific type of signal degradation as well as inter-talker differences, it does not allow us to definitively state whether one type of noise yields greater perceptual adaptation. Further research is required to explore this issue more fully.

The findings from the current study suggest that results from perceptual learning experiments using only one talker should be regarded with some caution, particularly with respect to the size of the perceptual learning effect. Most perceptual learning studies only use one talker or do not explicitly explore inter-talker differences. In line with our findings for degraded speech, Bradlow and Bent (2008) recently found that the extent of adaptation to foreign accented speech varied across talkers. Specifically, they found that the adaptation was greater for talkers with higher overall intelligibility. The issue of how perceptual learning is affected by overall intelligibility should be further explored with regard to the perception of degraded speech.

## B. Talker gender

Previous studies have reported that adult female talkers are more intelligible than adult male talkers for normal-hearing adult and child listeners both in quiet and in low levels of noise (Bradlow *et al.*, 1996; Hazan and Markham, 2004) and for Lombard speech (Junqua, 1993). This result has been consistently observed across different types of materials (i.e., both words and sentences). The findings from the current study are consistent with these previous results and support the claim that female talkers tend to be more intelligible than male talkers in tests of talker intelligibility. The present study adds to the previous findings by demonstrating that this result holds under two types of signal degradation (e.g., under CI-simulation and with speech mixed with multi-talker babble). However, it should be noted that the same talkers were used in the current experiment as in the study of Bradlow *et al.* (1996)

The source of the gender difference is not known at this point. It is possible that female talkers are generally more intelligible than their male counterparts because of physical differences in the vocal tracts. However, the gender differences could stem from a learned source of behavior. For example, female talkers could make more extreme articulatory adjustments that result in more intelligible speech at the segmental level. If this latter type of explanation is the source of this difference, it would suggest that male talkers

could possibly be taught to alter their articulatory patterns to increase their intelligibility. Furthermore, it remains possible that women produce speech differently than men when being recorded by adopting a clearer speaking style even when not explicitly instructed to. More work is needed to resolve this issue.

## V. CONCLUSIONS

The present results suggest that across-talker intelligibility differences are maintained under two types of signal degradation. High intelligibility talkers under CI-simulation also tended to be high intelligibility talkers in multi-talker babble listening conditions. These results replicate and extend the earlier intelligibility results of Green *et al.* (2007) by demonstrating that for a large number of talkers, intelligibility scores were significantly correlated for simulated CI listeners and normal-hearing listeners in noise. Furthermore, listeners were found to adapt rapidly to speech in both the CI-simulated and multi-talker babble conditions although greater perceptual adaptation was observed in the CI-simulation condition than in the multi-talker babble condition, and the extent of adaptation differed widely across talkers and listeners.

## ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health to Indiana University (NIH-NIDCD T32 Grant No. DC-00012 and NIH-NIDCD R01 Grant No. DC-000111). An earlier version of this work was presented at the fourth Joint Meeting of Acoustical Society of America and the Acoustical Society of Japan, Honolulu, HI, November 2006. We thank Wesley Alford, Vidhi Sanghavi, Melissa Troyer, and Jennifer Karpicke for their assistance in data collection, Luis Hernandez for technical assistance, Larry Phillips for help with data entry, Ann Bradlow for allowing us access to her data, Jeremy Loebach for being so generous with his time, and Rochelle Newman and two anonymous reviewers for their many helpful suggestions.

<sup>1</sup>In this paper, we operationally define speech intelligibility as the listener's ability to accurately report the words that a talker has produced. This objective measure of speech intelligibility contrasts with other measures in which listeners subjectively rate the "intelligibility" of a speaker (also called comprehensibility; e.g., Fayer and Krasinski, 1987) or tests in which the listener must provide an accurate paraphrase of the talker's message for the talker's communicative intent to be considered effective (e.g., Brodkey, 1972).

Assman, P. F., and Summerfield, A. Q. (2004). "The perception of speech under adverse conditions," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. Fay (Springer-Verlag, New York).

Bent, T., and Bradlow, A. R. (2003). "The interlanguage speech intelligibility benefit," *J. Acoust. Soc. Am.* **114**, 1600–1610.

Black, J. W. (1957). "Multiple-choice intelligibility tests," *J. Speech Hear. Disord.* **22**, 213–235.

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

Bradlow, A. R., and Bent, T. (2008). "Perceptual adaptation to non-native speech," *Cognition* **106**, 707–729.

Bradlow, A. R., Toretta, G. M., and Pisoni, D. B. (1996). "Intelligibility of

normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," *Speech Commun.* **20**, 255–272.

Brodkey, D. (1972). "Dictation as a measure of mutual intelligibility: A pilot study," *Lang. Learn.* **22**, 203–220.

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

Clopper, C. G., and Bradlow, A. R. (2008). "Perception of dialect variation in noise: Intelligibility and classification," *Lang. Speech* **51**, 175–198.

Cox, R. M., Alexander, G. C., and Gilmore, C. (1987). "Intelligibility of average talkers in typical listening environments," *J. Acoust. Soc. Am.* **81**, 1598–1608.

Davis, M. H., Johnsrude, I. S., Hervais-Ademan, A., Taylor, K., and McGestigan, 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.

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.

Fayer, J. M., and Krasinski, E. (1987). "Native and non-native judgments of intelligibility and irritation," *Lang. Learn.* **37**, 313–326.

Felty, R., Buchwald, A., and Pisoni, D. B. (2009). "Adaptation to frozen babble in spoken word recognition," *J. Acoust. Soc. Am.* **125**(3), EL93–EL97.

Ferguson, S. H. (2004). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," *J. Acoust. Soc. Am.* **116**, 2365–2373.

Fletcher, H., and Steinberg, J. C. (1924). "The dependence of the loudness of a complex sound upon the energy in the various frequency regions of the sound," *Phys. Rev.* **24**, 306–318.

Francis, A. L., and Nusbaum, H. C. (2002). "Selective attention and the acquisition of new phonetic categories," *J. Exp. Psychol. Hum. Percept. Perform.* **28**, 349–366.

Francis, A. L., Nusbaum, H. C., and Fenn, K. (2007). "Effects of training on the acoustic phonetic representation of synthetic speech," *J. Speech Lang. Hear. Res.* **50**, 1445–1465.

Gonzalez, 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.

Green, T., Katiri, S., Faulkner, A., and Rosen, S. (2007). "Talker intelligibility differences in cochlear implant listeners," *J. Acoust. Soc. Am.* **121**(6), EL223–EL229.

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.

Hazan, V., and Markham, D. (2004). "Acoustic-phonetic correlates of talker intelligibility in adults and children," *J. Acoust. Soc. Am.* **116**, 3108–3118.

Healy, E. W., and Montgomery, A. A. (2007). "The consistency of sentence intelligibility across three types of signal distortion," *J. Speech Lang. Hear. Res.* **50**, 270–282.

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., and Carlyon, R. P. (2008). "Perceptual learning of noise vocoded words: Effects of feedback and lexicality," *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 460–474.

Hood, J. D., and Poole, J. P. (1980). "Influence of the speaker and other factors affecting speech intelligibility," *Audiology* **19**, 434–455.

Howes, D. (1952). "The intelligibility of spoken messages," *J. Psychol.* **65**, 460–465.

Howes, D. (1957). "On the relation between the intelligibility and frequency of occurrence of English words," *J. Acoust. Soc. Am.* **29**, 296–305.

IEEE (1969). "IEEE recommended practices for speech quality measurements," *IEEE Trans. Audio Electroacoust.* **17**, 227–246.

Imai, S., Flege, J. E., and Walley, A. (2003). "Spoken word recognition of accented and unaccented speech: Lexical factors affecting native and non-native listeners," in *Proceedings of the International Congress on Phonetic Science*, Barcelona, Spain.

Junqua, J.-C. (1993). "The Lombard reflex and its role on human listeners

- and automatic speech recognizers," *J. Acoust. Soc. Am.* **93**, 510–524.
- Karl, J., and Pisoni, D. B. (1994). "The role of talker-specific information in memory for spoken sentence," *J. Acoust. Soc. Am.* **95**, 2873.
- Kraljic, T., and Samuel, A. G. (2006). "How general is perceptual learning for speech?," *Psychon. Bull. Rev.* **13**, 262–268.
- Kraljic, T., and Samuel, A. G. (2007). "Perceptual adjustments to multiple speakers," *J. Mem. Lang.* **56**, 1–15.
- Labov, W., and Ash, S. (1997). "Understanding Birmingham," in *Language Variety in the South Revisited*, edited by C. Bernstein, T. Nunnally, and R. Sabino (University of Alabama Press, Tuscaloosa, AL).
- Lane, H., and Tranel, B. (1971). "The Lombard sign and the role of hearing in speech," *J. Speech Hear. Res.* **14**, 677–709.
- Lane, H., Tranel, B., and Sisson, C. (1970). "Regulation of voice communication by sensory dynamics," *J. Acoust. Soc. Am.* **47**, 618–624.
- Loebach, J. L., and Pisoni, D. B. (2008). "Perceptual learning of spectrally degraded speech and environmental sounds," *J. Acoust. Soc. Am.* **123**, 1126–1139.
- Lombard, E. (1911). "Le signe de l'elevation de la voix (The sign of elevating the voice)," *Annales de Maladies d L'oreille et du Larynx* **37**, 101–119.
- Major, R., Fitzmaurice, S., Bunta, F., and Balasubramanian, C. (2002). "The effects of nonnative accents on listening comprehension: Implications for ESL assessment," *TESOL Quarterly* **36**, 173–190.
- Mason, H. M. (1946). "Understandability of speech in noise as affected by region of origin of speaker and listener," *Speech Monographs* **13**, 54–68.
- McGarr, N. S. (1983). "The intelligibility of deaf speech to experienced and inexperienced listeners," *J. Speech Hear. Res.* **26**, 451–458.
- Miller, G. A. (1947). "The masking of speech," *Psychol. Bull.* **44**, 105–129.
- Miller, G. A., and Nicely, P. E. (1955). "An analysis of perception confusions among some English consonants," *J. Acoust. Soc. Am.* **27**, 338–352.
- Moore, B. C. J. (2003). "Speech processing for the hearing-impaired: Successes, failures and implication for speech mechanisms," *Speech Commun.* **41**, 81–91.
- Munro, M., Derwing, T., and Morton, S. (2006). "The mutual intelligibility of foreign accents," *Stud. Second Lang. Acquis.* **28**, 111–131.
- Norris, D., McQueen, J. M., and Cutler, A. (2003). "Perceptual learning in speech," *Cogn. Psychol.* **47**, 204–238.
- Nygaard, L. C., and Pisoni, D. B. (1998). "Talker-specific learning in speech perception," *Percept. Psychophys.* **60**, 335–376.
- Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (1994). "Speech perception as a talker-contingent process," *Psychol. Sci.* **5**, 42–46.
- Pallier, C., Sebastian-Gallés, N., Dupoux, E., Christophe, A., and Mehler, J. (1998). "Perceptual adjustment to time-compressed speech: A cross-linguistic study," *Mem. Cognit.* **26**, 844–851.
- Picheny, M. A., Durlach, N. I., and Braida, L. D. (1985). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," *J. Acoust. Soc. Am.* **28**, 96–103.
- Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., and Abrams, H. B. (2006). "Effects of bilingualism, noise and reverberation on speech perception by listeners with normal hearing," *Appl. Psycholinguist.* **27**, 465–485.
- 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., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (1988). "Effects of noise on speech production: Acoustic and perceptual analyses," *J. Acoust. Soc. Am.* **84**, 917–928.
- Uchanski, R. M., Choi, S., Braida, L. D., Reed, C. M., and Durlach, N. I. (1996). "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate," *J. Speech Hear. Res.* **39**, 494–509.
- van Wijngaarden, S. J. (2001). "Intelligibility of native and non-native Dutch speech," *Speech Commun.* **35**, 103–113.
- van Wijngaarden, S. J., Steeneken, H. J. M., and Houtgast, T. (2002). "Quantifying the intelligibility of speech in noise for non-native listeners," *J. Acoust. Soc. Am.* **111**, 1906–1916.
- Weil, S. A. (2001). "Foreign-accented speech: Encoding and generalization," *J. Acoust. Soc. Am.* **109**, 2473.