RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 24 (2000)
Indiana University

Effects of Speaking Style on the Perceptual Learning of Novel Voices: A First Report[1]

James D. Harnsberger, Richard Wright, and David B. Pisoni

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

[1] This work was supported by NIH-NIDCD Training Grant DC00012 and NIH-NIDCD Research Grant R01-DC00111 to Indiana University. We would like to thank Kipp McMichael for his technical assistance.

Effects of Speaking Style on the Perceptual Learning of Novel Voices: A First Report

Abstract. This study examined the effects of speaking style on the perceptual learning of novel voices in the laboratory. Listeners participated in a voice learning experiment. In the training phase, listeners were asked to learn the names of either seven male or seven female talkers from samples of citation or hyperarticulated speech. In the test phase, listeners were presented with the same stimuli as in the training phase and were asked to identify the talker, with no feedback. In the sentence generalization phase, listeners were asked to identify the same voices producing new sentences in the same speaking style as that used in the previous phases. In the speaking style generalization phase, listeners were asked to identify the same voices in either the same speaking style as the previous phases or in a novel speaking style. The results showed that female voices were easier to learn in a hyperarticulated speaking style relative to a citation speaking style in the training and test phases. For the male voices, no such effect was observed. In addition, voice identification scores increased from the training to the test phase. However, voice identification scores did not improve in subsequent phases, which lacked the feedback provided during the training phase. In the style generalization phase, training with the female voice hyperarticulated tokens provided a greater advantage in identifying voices in a novel style relative to training with female voice citation tokens. No such effect was observed for listeners trained with the male voices. This gender interaction was further explored in a similarity scaling experiment, using stimuli from the first experiment. Listeners were presented with pairs of stimuli that differed in talker but matched in speaking style and gender. The citation sentences of both male and female talkers were rated as significantly more similar than the hyperarticulated sentences. However, for a subset of the stimuli, the difference in mean similarity for the female citation and hyperarticulated sentences was significantly greater than the corresponding difference in the male voices, indicating that the female talkers may have produced a more perceptually distinct hyperarticulated style than the male talkers. These differences may have contributed to the gender effect observed in learning in the first experiment. Taken together, the results of both experiments show that speaking style exerts a strong influence on the learning of novel voices, but its exact role is unclear given the interaction of speaking style and gender of the talker.

Introduction

Prior work on voice perception has shown that numerous sources of variability in the stimulus materials and in the listeners tested affect the encoding and retention of voice information in long-term memory (Kreiman, 1997). In terms of stimulus characteristics, voice perception has been shown to be influenced by the phonetic characteristics of the stimulus materials (see Bricker & Pruzansky, 1976 for a review), the filtering of glottal or vocal tract characteristics (Kubawara & Takagi, 1991; Tartter, 1991), and changes in the speaking rate or any signal distortion (Van Lancker, Kreiman, & Wickens, 1985). Several characteristics of listener groups have also been manipulated. For example, mismatches between the native dialect or language of the listeners and that of the talkers who produced the utterances have resulted in poorer voice recognition performance relative to controls that match in terms of linguistic background (Hollien, Majewski, & Doherty, 1982; Thompson, 1987; Goggin, Thompson, Strube, & Simental, 1991). Thus, linguistic information influences the voice recognition process, much like talker information has been shown to affect speech perception and spoken word recognition (Nygaard & Pisoni, 1998).

Individual listeners have been shown to vary widely in the specific acoustic cues they use to judge the relative similarity of voices (Hollien, Majewski, & Doherty, 1982; Kreiman, Gerratt, Precoda, & Berke, 1992). Listeners also vary in the strategies used to identify novel male and female voices that potentially reflect what some researchers have referred to as "cultural stereotypes" concerning male and female voices. Examples of such stereotypes include the association of "breathiness" with female voices and "hoarseness" with male voices (Singh & Murray, 1978), or associations between different distributions of vowel categories and gender that cannot be explained solely by differences in vocal tract size (Mattingly, 1966). In examining the voice perception literature overall, it is clear that the attributes of the voices used in experiments and the attributes associated with listeners' experience with talkers in their native language both play significant roles in the discrimination and recognition of novel voices.

While several sources of variability have been investigated in voice perception research, an important source of variation in speech production has been neglected in these earlier studies. Specifically, variation in speaking style has not been studied in any great depth. Listeners on a daily basis not only encounter a variety of talkers, including ones whose voices are unfamiliar to the listener, but they are also exposed to a wide variety of speaking styles. Speaking styles can vary along a continuum from casual to careful speech (Labov, 1970). Moreover, speaking styles are specific to a given situation or goal (e.g., performance speech; speech directed to authority figures; speech in noisy environments). Variation in speaking style results in significant changes in the acoustic characteristics of the speech signal that are linguistically relevant. This variation also provides information about the identity of the talker, through such cues as f0, formant frequencies, nasality, duration, and breathiness (Murray & Singh, 1980). Thus, it is possible that information about speaking style variation is encoded with voice information and, therefore, may influence the process of learning novel voices.

The goal of this study was to examine the effects of speaking style on the learning of novel voices. The speaking styles used in this study were *citation* speech, corresponding to read speech commonly elicited in the laboratory, and *hyperarticulated* speech, a style involving a high degree of articulatory precision produced by a talker who is attempting to speak clearly. Samples of citation style speech were obtained by simply having talkers read aloud sentences from a computer screen. Samples of hyperarticulated speech were elicited via the methods developed recently by Brink, Wright, and Pisoni (1998), in which talkers were asked to repeat a sentence more clearly after having produced it once before in a citation style. These two speaking styles were selected because in earlier pilot studies with utterances from five female voices, a significant increase was observed in the rate of learning of female voices from citation sentences compared to hyperarticulated sentences.

In the present study, we sought to extend this work to male voices, and to additional learning tasks. Specifically, we included a sentence generalization task (same voices, same style, novel sentences) and a speaking style generalization task (same voices, novel style, novel sentences). The sentence generalization task was added to determine whether the effects of speaking style were specific to the stimulus materials used in the training and test phases. The speaking style generalization task was administered to determine if either the citation or the hyperarticulated speaking style conveyed the most information that could be used by listeners in learning voices in novel styles.

For this study, two hypotheses were assessed concerning the effects of speaking style on novel voice learning. First, voices would be easier to learn in the hyperarticulated speaking style relative to the citation speaking style. This prediction was based on the pilot results, as well as prior research in speech perception and spoken word recognition in which hyperarticulated speech is typically found to be more intelligible and more informative in the identity of words than more reduced styles, including citation speech (Lindblom, 1990; Moon & Lindblom, 1994; Picheny, Durlach, & Braida, 1985; Picheny, Durlach,

& Braida, 1986). This hypothesis was termed the *clear speech hypothesis*. An alternative prediction was also entertained, that voices would be easier to learn in a citation rather than a hyperarticulated style. Voices in a citation speech style were hypothesized to be more distinctive because, under more casual speaking styles, idiosyncratic gestural strategies emerge that are otherwise suppressed by the use of stereotyped articulatory gestures in hyperarticulated speech. Such idiosyncrasies could be the product of the vocal anatomy of the talker, or could simply be an individual strategy to reach a common acoustic/auditory target (Johnson, Ladefoged, & Lindau, 1993). In either case, talker gestural idiosyncrasies could be available to the listener and constitute useful cues in learning to identify individual talkers. This hypothesis was termed the *idiosyncratic articulation hypothesis*.

Both hypotheses were tested in a voice learning task in which subjects were trained to identify novel male or female voices (Experiment 1). Subjects were trained on a set of sentences from each talker and tested on the same sentences. Subjects were also asked to identify the same voices producing novel sentences and, in some conditions, novel speaking styles. The role of speaking style in voice learning was further examined in a voice similarity task (Experiment 2), in which subjects were asked to rate on a seven-point scale the relative similarity of novel male and female voices producing sentences in a citation and hyperarticulated style. The voice similarity task was designed to measure the relative similarity of male and female voices in different speaking styles in an effort to account for patterns observed in the voice learning experiment.

EXPERIMENT 1

Methods

Participants

One hundred and six native speakers of American English, 52 females and 54 males ranging in age from 18 to 23 ($M = 20$), participated in this study. Participants received course credit for participating in a single one-hour test session. None of the listeners reported any history of a speech or hearing disorder at the time of testing.

Stimulus Materials

Recordings of sentences from 14 native speakers of American English, seven females and seven males ranging in age between 18 and 30, were used in this study. Participants used for creating the stimulus materials received $15 total compensation for participating in two one-hour sessions. None of the participants reported any history of a speech or hearing disorder at the time of testing. Participants were recorded reading 34 sentences chosen from the 200 sentences comprising the SPIN set (Kalikow, Stevens, & Elliot, 1977). The SPIN sentences are short sentences, five to eight words in length, ending in a high frequency monosyllabic noun (e.g., “The farmer harvested his crop”). The 34 SPIN sentences selected for this study are listed in Appendix A. The recording session took place in a sound-attenuated chamber (IAC Audiometric Testing Room, Model 402) using a head-mounted Shure (SM98) microphone positioned one inch away from the subject's chin. The recordings were digitized at 22.05 kHz (16 bit sampling) using a Tucker-Davis Technologies System II and stored on an IBM-PC 486 computer. The 34 sentences were produced using three speaking styles via a method used by Harnsberger and Pisoni (1999), a modified version of a method first developed by Brink, Wright, and Pisoni (1998). Those speaking styles were reduced (i.e., casual, hypoarticulated), citation (i.e., read speech), and hyperarticulated (i.e., careful speech). Of the 34 sentences in these three styles, 15 sentences in two styles, citation and hyperarticulated, were used in the present voice learning study. The fifteen sentences presented across all four phases appear in Appendix B.

Procedures

The participants for the voice learning experiment were randomly assigned to one of eight experimental conditions. Each condition consisted of four phases, in the following order: (1) a Training phase, (2) a Test phase, (3) a Sentence Generalization phase, and (4) a Speaking Style Generalization phase. The eight conditions differed in terms of the gender of the talkers whose stimuli were presented in the four phases (male or female), the speaking style of the stimuli presented in the first three phases, and the speaking style of the stimuli presented in the Speaking Style Generalization phase. All eight conditions are outlined in Table 1.

Gender	Style		Name of condition
	Phases 1 - 3	Phase 4	
Male	Citation	Citation	Male Cit-Cit
	Citation	Hyperarticulated	Male Cit-Hyp
	Hyperarticulated	Citation	Male Hyp-Cit
	Hyperarticulated	Hyperarticulated	Male Hyp-Hyp
Female	Citation	Citation	Female Cit-Cit
	Citation	Hyperarticulated	Female Cit-Hyp
	Hyperarticulated	Citation	Female Hyp-Cit
	Hyperarticulated	Hyperarticulated	Female Hyp-Hyp

Table 1. The eight conditions of Experiment 1.

In Table 1, *Gender* refers to the gender of the talkers who produced the stimuli in that condition. *Style* refers to the speaking style used by the talkers in that condition, in *Phases 1 - 3* of the experiment (i.e., Training, Test, and Sentence Generalization) and in *Phase 4* (Speaking Style Generalization). The *Name of condition* column lists the name assigned to each condition of the experiment. For example, listeners who learned male citation voices in all four phases of the experiment had been assigned to the "Male Cit-Cit" condition.

Training Phase. In this phase, participants were presented with a block of sentences and asked to identify the talker who produced each sentence. Two repetitions of five different sentences from each of the seven talkers in a given condition were presented in random order, for a total of 70 trials. Participants were asked to identify each voice by pressing one of seven buttons on a keyboard. Each button was labeled with a name (Ben, Greg, James, Kyle, Matt, Max, and Steve for the male voice sentences; Jenny, Kim, Lynn, Mary, Paula, Susie, and Trixie for the female voice sentences). After each response, the correct name of the talker appeared on the computer screen.

Test Phase. After participants had completed the Training phase, they were presented with one repetition of the same 35 sentences used in the Training phase. Participants were asked to identify each voice by pressing the appropriate button on the keyboard. In this phase, no feedback was provided to the participants.

Sentence Generalization Phase. Following the Test phase, participants were presented with five new sentences from the same seven talkers who produced the stimuli used in the Training and Test phases. As before, participants were asked to identify each voice by pressing the appropriate button on the keyboard. No feedback was provided.

Speaking Style Generalization Phase. Following the Sentence Generalization phase, participants were presented with five new sentences from the same seven talkers who produced the stimuli used in the previous three phases. The sentences used in this condition were produced in either the same speaking style used in the previous phases or in a novel speaking style, depending on which condition the individual participants were assigned to. Participants were asked to identify each voice by pressing the appropriate button on the keyboard. No feedback was provided.

Predictions

Two hypotheses concerning voice learning were entertained in this study: the *clear speech hypothesis* and the *idiosyncratic articulation hypothesis*. According to the *clear speech hypothesis*, voices should be easier to learn in a hyperarticulated style of speech than in a citation style because hyperarticulated sentences are produced with the more extreme articulatory gestures, resulting in longer segments, lengthy pauses between words, and an expanded vowel space. Thus, we expected significantly higher identification scores in the hyperarticulated speech conditions (Hyp-Cit, Hyp-Hyp) than in the citation speech conditions (Cit-Cit, Cit-Hyp) in the first three phases of the experiment. In the Speaking Style Generalization phase, experience with hyperarticulated speech was predicted to transfer more easily to learning voices in a novel style than experience with citation speech. Thus, the mean identification scores for participants in the Hyp-Cit conditions were predicted to be significantly higher than those of the participants in the Cit-Hyp conditions.

According to the *idiosyncratic articulation hypothesis*, less monitored and more naturally produced speaking styles should display talker-specific idiosyncratic speech production strategies that may provide more diverse cues to the identity of the talker. In this case, we expected to see significantly higher identification scores in the citation speech conditions relative to the hyperarticulated speech conditions. Moreover, in the Speaking Style Generalization phase, participants who learned to identify voices from citation speech should display an advantage in identifying voices in a novel speaking style relative to participants who learned from hyperarticulated speech. Thus, the mean identification scores for participants in the Cit-Hyp conditions should be significantly higher than those of the participants in the Hyp-Cit conditions.

Results

Training Phase

The proportion of correct responses to stimuli in the training phase of the experiment appear in Figure 1, listed by four conditions (Male Citation, Male Hyperarticulated, Female Citation, Female Hyperarticulated) rather than the full eight conditions due the lack of a significant three-way interaction between Gender, Speaking Style, and Speaking Style Generalization (see below). The four conditions were calculated by averaging together results from conditions that shared a common Gender and Speaking Style in the first three phases of the experiment (see Table 1). Thus, the results from the Female Cit-Cit and Female Cit-Hyp groups were combined in a Female Citation mean score. In a similar manner, Female Hyp-Cit and Female Hyp-Hyp; Male Cit-Cit and Male Cit-Hyp; and Male Hyp-Cit and Male Hyp-Hyp were combined in Female Hyperarticulated, Male Citation, and Male Hyperarticulated scores, respectively.

The results show an unanticipated effect of the gender of the talkers on voice learning across different speaking styles. Listeners who were assigned to the female voice hyperarticulated stimuli were more successful on average in learning new voices than the listeners assigned to the female voice citation stimuli. In contrast, listeners assigned to the male voice citation stimuli were more successful in learning new voices than the listeners assigned to the male voice hyperarticulated stimuli.

The results of the training phase of the experiment were submitted to an ANOVA with three between-subjects factors: Gender (male, female); Speaking Style (citation, hyperarticulated); and Speaking Style Generalization, which refers to the stimulus set presented in the Speaking Style Generalization condition (citation, hyperarticulated). None of the main effects were significant. Of the two- and three-way interactions, only the Gender by Speaking Style interaction was significant (F (1, 98) = 4.97, $p < .05$). A simple effects analysis showed that the effect of Speaking Style was significant within the female voice sets but not the male voice sets (Female: F (1, 98) = 5.65, $p < .05$; Male: F (1, 98) = 0.7, $p = .4$). A simple effects analysis of Gender within Speaking Style showed no significant differences.

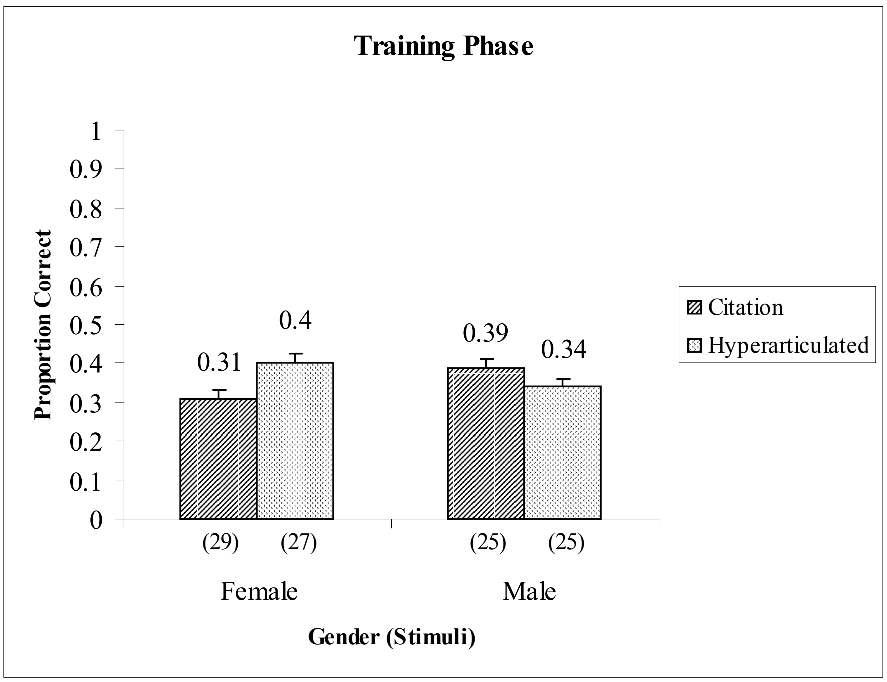

Figure 1. The mean proportion correct responses in the training phase for the four groups assigned to four different stimulus sets (Female: Citation, Hyperarticulated; Male: Citation, Hyperarticulated). Values in parentheses below each column denote the number of listeners in each group. Error bars denote the standard error of the mean.

Test Phase

Figure 2 shows the results from the test phase of the experiment. As in the training phase, the three-way interaction between Gender, Speaking Style, and Speaking Style Generalization was not significant. Once again, the results are reported in Figure 2 in four conditions, averaging over differences in Speaking Style Generalization. As in the training phase, a gender effect was observed in the identification scores for the test phase. Listeners who were assigned to the female voice hyperarticulated stimuli were more successful on average in learning new voices than the listeners assigned to the female voice citation stimuli. In contrast, listeners assigned to the male voice citation stimuli were more successful in learning new voices than the listeners assigned to the male voice hyperarticulated stimuli.

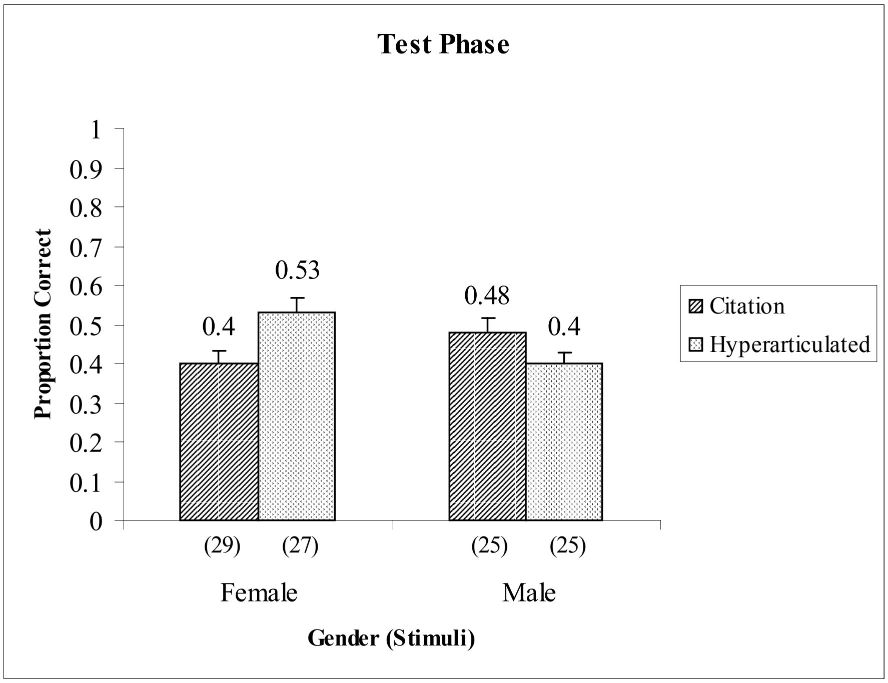

Figure 2. The mean proportion correct responses in the test phase for the four groups assigned to four different stimulus sets (Female: Citation, Hyperarticulated; Male: Citation, Hyperarticulated). Values in parentheses below each column denote the number of listeners in each group. Error bars denote the standard error of the mean.

The results of the test phase of the experiment were submitted to an ANOVA with three between-subjects factors: Gender (male, female), Speaking Style (citation, hyperarticulated), and Speaking Style Generalization (citation, hyperarticulated). None of the main effects were significant. Of the two-way interactions, only the Gender by Speaking Style interaction was significant (F (1, 98) = 8.47, $p < .01$). A simple effects analysis showed that the effect of Speaking Style was significant within the female voice sets but only marginally significant in the male voice sets (Female: F (1, 98) = 5.68, $p < .05$; Male: F (1, 98) = 3.09, $p = .08$). A simple effects analysis of Gender within Speaking Style also showed that the effect of Gender was significant within the hyperarticulated style sets but not the citation style sets (Hyperarticulated: F (1, 98) = 6.22, $p < .05$; Citation: F (1, 98) = 2.61, $p = .11$).

The results from the training and the test phases were also compared to determine if training resulted in significant improvement in the learning of novel voices. In a four-way ANOVA with Phase as a within-subject factor (training, test) and Gender, Speaking Style, and Speaking Style Generalization as between-subject factors, the main effect of Phase (F (1, 98) = 66.23, $p < .001$) and the interaction of Phase by Gender by Speaking Style (F (1, 98) = 4.29, $p < .05$) were significant. No other main effects or interactions were significant.

The significant effect of Phase showed that listeners did improve overall in the ability to identify novel voices. The mean proportion correct scores for the training and test phases were 0.36 and 0.46, respectively. The significant three-way interaction was further explored using simple effects analyses. The results of these analyses showed that all four listener groups (Female Citation, Female Hyperarticulated, Male Citation, and Male Hyperarticulated) improved significantly ($p < .01$ in all cases) from the training phase to the test phase.

Sentence Generalization Phase

Figure 3 displays the results of the sentence generalization phase of the experiment. As in the previous phases, listeners in the Female Hyperarticulated condition achieved higher identification scores than listeners in the Female Citation condition. The pattern for the male voice conditions showed the opposite effect of gender and speaking style. In a three-way ANOVA, none of the main effects proved to be significant. Of the interactions, only the Gender by Speaking Style interaction was significant (F (1, 98) = 4.01, $p < .05$). Given that the three-way interaction was not significant, the results in Figure 3 combine the groups that differ only in Speaking Style Generalization. The significant two-way interaction was assessed using simple effects analyses. Unlike the results of the previous phases, none of the tests of Speaking Style within Gender or Gender within Speaking Style were significant, although the difference between Female Hyperarticulated and Male Hyperarticulated approached significance (F (1, 98) = 3.57, $p = .06$).

The results from the sentence generalization and the test phases were also compared to determine if the performance of the subjects improved with additional exposure to the voices, without the feedback available in the training phase. The results of a four-way ANOVA with Phase as a within-subject factor (sentence generalization, test) and Gender, Speaking Style, and Speaking Style Generalization as between-subject factors showed that none of the main effects or interactions were significant. The results of this analysis revealed that the listener groups plateaued in their capacity to identify novel voices. Without additional feedback, simple exposure to the stimuli in the sentence generalization trials did not improve voice learning.

Speaking Style Generalization Phase

Figure 4 shows the results of the speaking style generalization phase of the experiment. The results are displayed in terms of the eight groups to which the listeners were assigned (Female: Cit-Cit, Cit-Hyp, Hyp-Cit, Hyp-Hyp; Male: Cit-Cit, Cit-Hyp, Hyp-Cit, Hyp-Hyp). Overall, for both the male and female stimulus sets, identification scores were highest when listeners were presented with the same speaking style in the speaking style generalization phase that they were exposed to in the preceding phases (i.e., listeners in the Male and Female Cit-Cit and Hyp-Hyp conditions). When listeners were presented with stimuli in a speaking style that differed from that used in preceding phases, identification scores were lower. Moreover, these lower scores varied by condition. Listeners who had learned voices from Female Hyperarticulated stimuli had higher identification scores in a novel speaking style (i.e., Citation) than listeners who had learned voices from Female Citation stimuli and were tested with Female Hyperarticulated stimuli in the speaking style generalization phase. An opposite pattern was observed in male voice stimuli: listeners in the Male Cit-Hyp condition had slightly higher identification scores than listeners in the Male Hyp-Cit condition. Thus, the hyperarticulated speaking style was the most informative style for learning female voices and generalizing to voices in novel speaking styles, while the citation speaking style was the most informative for male voices.

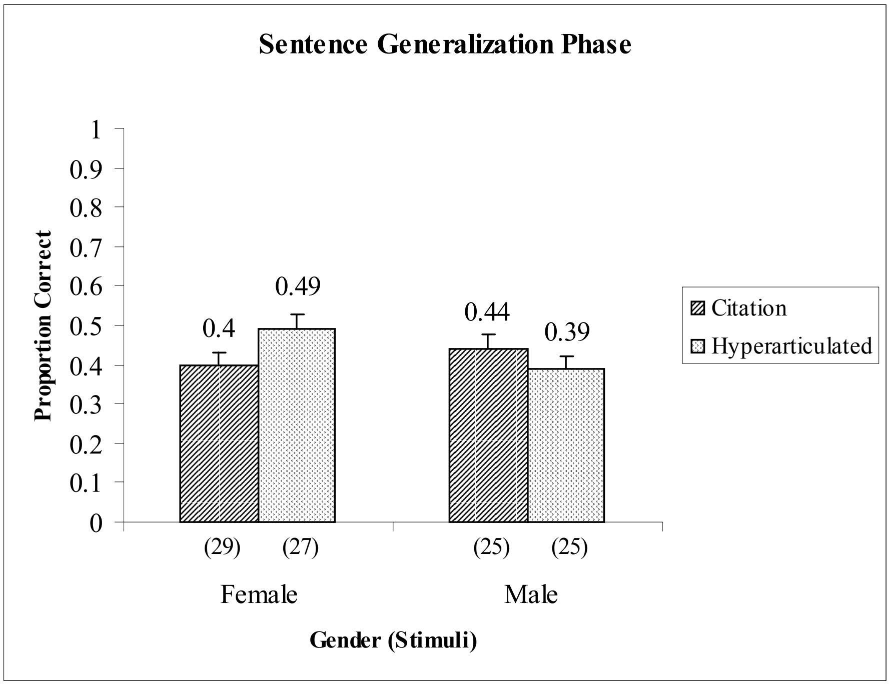

Figure 3. The mean proportion correct responses in the sentence generalization phase listed by four groups assigned to four different stimulus sets (Female: Citation, Hyperarticulated; Male: Citation, Hyperarticulated). Values in parentheses below each column denote the number of listeners in each group. Error bars denote the standard error of the mean.

A three-way ANOVA of the speaking style generalization results showed that none of the main effects of Gender, Style, or Speaking Style Generalization were significant. Of the two-way interactions, Gender by Style (F (1, 98) = 5.22, $p < .05$) and Style by Speaking Style Generalization (F (1, 98) = 10.9, $p < .001$) were significant. A simple effects analysis of the Gender by Style interaction showed that listeners who had learned female hyperarticulated voices in the previous phases were significantly better in identifying female voices than listeners who had learned female citation voices in the previous phases. Listeners who had learned female hyperarticulated voices in previous phases averaged 0.48 proportion correct on female voices in this phase, as compared with 0.40 proportion correct for listeners who learned female voices in a citation style in previous phases.

A simple effects analysis of the Speaking Style by Speaking Style Generalization interaction showed that, in several cases, performance in voice learning decreased when listeners were asked to identify voices in a novel speaking style. Listeners who learned citation voices in previous phases were significantly better in identifying citation voices in the speaking style generalization phase (i.e., Cit-Cit listeners) than listeners who learned hyperarticulated voices in the previous phases (Cit-Hyp listeners) (F (1, 98) = 4.9, $p < .05$). Cit-Cit and Cit-Hyp listeners averaged 0.49 and 0.35 proportion correct, respectively. Cit-Cit listeners also outperformed listeners who learned hyperarticulated voices and were presented with citation voices in this phase (Hyp-Cit listeners, who averaged 0.38 proportion correct) (F (1, 98) = 9.29, $p < .01$). Finally, Hyp-Hyp listeners identified voices significantly better than Hyp-Cit listeners, averaging 0.46 and 0.38 proportion correct, respectively (F (1, 98) = 6.05, $p < .05$).

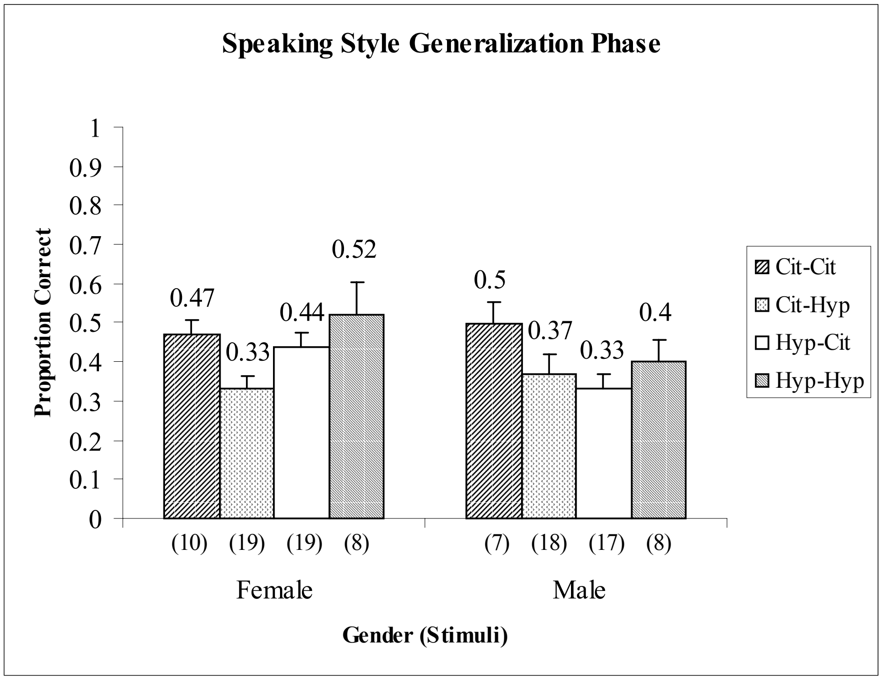

Figure 4. The mean proportion correct responses in the speaking style generalization phase listed by eight groups assigned to eight different stimulus sets (Female: Cit-Cit, Cit-Hyp, Hyp-Cit, Hyp-Hyp; Male: Cit-Cit, Cit-Hyp, Hyp-Cit, Hyp-Hyp). Values in parentheses below each column denote the number of listeners in each group. Error bars denote the standard error of the mean.

The results from the speaking style and sentence generalization phases were also compared to determine if the performance of the subjects improved with additional exposure to the voices. A four-way ANOVA with Phase as a within-subject factor (speaking style generalization, sentence generalization) and Gender, Style, and Speaking Style Generalization as between-subject factors showed only a significant effect of Phase (F (1, 98) = 5.44, p < .05). Performance on the speaking style generalization phase was significantly lower than performance in the sentence generalization phase, reflecting the poorer scores of listeners who identified voices in a novel speaking style.

Discussion

In the first experiment, we tested two hypotheses concerning the effects of speaking style on learning novel voices. One hypothesis predicted that participants would be more successful in learning voices from citation speech than hyperarticulated speech because of the absence of talker-specific gestural strategies that are suppressed when producing hyperarticulated speech. The second hypothesis made the opposite prediction. The results showed an unanticipated gender effect that makes evaluating hypotheses about speaking style difficult because the effects are influenced by gender.

With the seven female voices, the results in all training and test phases of the experiment supported the *clear speech hypothesis*. We found that it was easier to learn female voices in a hyperarticulated speaking style than female voices in a citation speaking style in the training and test phases. Learning from a female voices in a hyperarticulated speaking style also generalized more readily to identifying the same voices in a novel speaking style than learning from female voices in a citation speaking style. No significant effects of speaking style were found for the male voices, although there was a trend for male voices in a citation speaking style to be easier to learn than male voices in a hyperarticulated speaking style. In addition, for the listeners who learned female voices, exposure to hyperarticulated voices in the training, test, and sentence generalization phases improved performance in identifying voices in a novel speech style relative to exposure to citation voices. Finally, across the four phases of the experiment, learning in general was enhanced by the feedback available in the training phase, but did not increase in subsequent phases.

In interpreting the results of the first experiment, it should be noted that some of the marginally significant or nonsignificant results with both the male and female voice sets might be a function of relatively low power. In the Female Cit-Cit, Female Hyp-Hyp, Male Cit-Cit, and Male Hyp-Hyp conditions, only seven to ten listeners were tested. In the original design of the experiment, we did not anticipate that gender would play a role in novel voice learning and we planned to combine data from both gender conditions. Thus, the results from the matching Female and Male conditions (e.g., Female Cit-Cit and Male Cit-Cit) were to be combined into analyses of style-based conditions (e.g. Cit-Cit, Cit-Hyp, Hyp-Cit, Hyp-Hyp). Using this approach, an adequate number of listeners were tested in the Female Cit-Cit, Female Hyp-Hyp, Male Cit-Cit, and Male Hyp-Hyp conditions. However, given the gender effect that was observed, a larger number of participants should be recruited in future studies in all conditions, to ensure that the lack of results is not due to inadequate power.

The results of the experiment demonstrate an effect of speaking style on learning new voices, although the effect of gender makes interpreting the results more complex. In formulating the two hypotheses, we assumed that in either the citation or the hyperarticulated styles, greater talker-specific detail would be available for learners to encode. According to the *idiosyncratic articulation hypothesis*, talkers in a less careful, less monitored speaking style would produce more idiosyncratic gestural strategies and show greater overall articulatory/acoustic variability, all of which would constitute important and useful cues in distinguishing different voices. According to the *clear speech hypothesis*, hyperarticulated speech constituted a clear, information-rich signal that delineates the entire gestural space used by the individual talker.

Both hypotheses assumed that the attributes of one of the two styles would provide more information for voice learning. With the gender effect observed here, if the attributes of the speaking style account for the ease of learning female voices in a hyperarticulated speaking style, then the male and female participants may have adopted quite different strategies when prompted to produce the two speaking styles. If the two gender groups adopted somewhat different strategies in producing the two speaking styles, then presumably we would see a gender effect on the perceptual judgments of phonetically trained and naïve listeners with these two speaking styles (see Harnsberger & Pisoni, 1999) and on the acoustic analysis of sentences produced in these speaking styles (see Harnsberger & Goshert, 2000). In fact, no gender effect was observed in either of these studies. In particular, Harnsberger and Goshert (2000) found consistent differences across all talkers regardless of gender between male and female talkers' citation and hyperarticulated speech. Citation speech, on average, had shorter keyword durations and shorter sentence durations than Hyperarticulated speech. Moreover, vowels in keywords produced in a Citation style were more centralized, resulting in a smaller vowel space than vowels in keywords produced in a Hyperarticulated style, for all talkers regardless of gender. The male and female talkers may differ in their production of the two styles in dimensions not measured by Harnsberger and Goshert, although the link between such dimensions and the rate of learning novel voices is unclear.

While an explanation for the gender effect focusing on the attributes of the stimuli would be the most satisfactory one, there are other possible explanations that focus on the listeners' prior experience with male and female voices, and how such experience might influence novel voice learning. The gender effect revealed here is reminiscent of a frequently reported dichotomy in male and female speaking styles.

Specifically, the characteristics associated with hyperarticulated speech (expanded vowel space, maintenance of consonant clusters) appear to occur more frequently in female speech, while male speech typically shows more instances of "reduction" phenomena, such as a compressed vowel space and consonant cluster reduction/deletion (Fischer, 1958; Trudgill, 1974; Byrd, 1994; Bradlow, Toretta, & Pisoni, 1996). This gender-based dichotomy has been observed in several acoustic-phonetic studies of "laboratory speech," sociolinguistic studies (using impressionistic coding), and in transcriptions of spontaneous speech corpora. For instance, Byrd (1994) examined gender-based and regional dialect patterns in the TIMIT database, a large sentence database incorporating 630 talkers who read a total of 2,342 sentences. She found gender differences in speaking rate; the release of sentence-final stops; and the frequency of occurrence of schwa, glottal stops, syllabic [n], voiceless schwa, and [h]. In all of these gender differences, male speech displayed the more reduced forms (e.g., faster speaking rate, less frequent release of final stops, more likely to use schwa).[2] Bradlow, Toretta, and Pisoni (1996) measured several acoustic-phonetic characteristics in their study of the correlates of intelligible speech in a multitalker database. They observed that female speech was significantly more intelligible than male speech, and that female speech and male speech differed in some, but not all, of the attributes associated with careful, clear speech, such as fundamental frequency range or the timing relationship between adjacent segments. Several studies have shown that women use phonological forms associated with more "standard" speech more frequently than men (Fischer, 1958; Trudgill, 1974).

Overall, along a continuum of casual to careful speech, the typical speaking style employed by men may be found closer to the casual end of the continuum than the typical speaking style employed by women. This trend may be related to the gender effect observed in this study. Our prior experience with male and female voices in different speaking styles may influence the encoding of novel male and female voices in long-term memory which, in turn, influences the learning of novel voices. In other words, listeners possess a greater familiarity with male reduced speech and female hyperarticulated speech than their corresponding counterparts, male hyperarticulated speech and female reduced speech. This familiarity implies that listeners may have a bias in extracting cues to identify individual male and female talkers in different styles. To account for the results of this study, this bias would have to be rather robust, and would presumably operate under a variety of noisy conditions as well as the clear condition in which the stimuli were presented in this study. To date, such a robust effect of prior experience with male and female voices has not been documented in other studies. However, this effect might be an example of a "cultural stereotype" that listeners have acquired for male and female speech (Singh & Murray, 1978).

Given the unexpected gender effect observed in Experiment 1, the role of speaking style and gender in the learning of novel voices was explored in a second study. The purpose of the second experiment was to test two possible explanations for the gender effect. First, the gender effect may reflect acoustic properties of the stimuli not measured by Harnsberger and Goshert (2000). The female and male talkers who produced the stimuli for Experiment 1 may have adopted somewhat different strategies in differentiating the citation from the hyperarticulated style, resulting in two stimulus sets (male and female voices) that differed in their relative similarity, and thus learnability. Alternatively, the gender effect may reflect the frequency with which listeners have encountered relatively "careful" female speech versus relatively "reduced" male speech. These two hypotheses can be distinguished by comparing male and female citation and hyperarticulated stimuli using perceptual similarity tests. In these tests, naïve listeners are asked to rate pairs of sentences in terms of their similarity. It is possible that male voices in a citation style are judged as equally similar to one another as male voices in a hyperarticulated style. In contrast, it is possible that female voices in a hyperarticulated style are judged to be less similar to one another than female voices in a citation style. Findings such as these would indicate that specific acoustic-phonetic

[2] Male and female talkers did not significantly differ in one phonetic process associated with reduced speech, specifically the palatalization of alveolar obstruents.

properties of the sentences not measured by Harnsberger and Goshert (2000) may be responsible for the gender effect observed in the voice learning experiment. Alternatively, if the results of such a similarity test fail to show any gender effects (as predicted by the acoustic analysis), then the patterns observed in the voice learning experiment may reflect prior experience with male and female speaking styles. To assess these hypotheses, we administered a scaling experiment to obtain measures of perceptual similarity for four sets of stimulus materials: female citation, female hyperarticulated, male citation, and male hyperarticulated speech

EXPERIMENT 2

Methods

Participants

Thirty-two native speakers of American English, 24 females and eight males ranging in age from 18 to 29 (M = 20), participated in this study. The listeners received either course credit or five dollars for participating in a single one-hour test session. None of the listeners reported any history of a speech or hearing disorder at the time of testing.

Stimulus Materials and Procedures

A subset of the stimulus materials used in Experiment 1 was presented to participants in this experiment. Specifically, the Citation and Hyperarticulated readings of two sentences, “The beer drinkers raised their mugs” and “I made the phone call from a booth,” were selected. The seven male and seven female talkers’ Citation and Hyperarticulated readings of these sentences were used, for a total of 28 different stimuli per sentence. Participants were divided randomly into two groups of 16 listeners each. The two groups were presented with tokens of either the first or the second sentence. The 28 stimuli represented four sets of stimulus materials: Male Citation, Female Citation, Male Hyperarticulated, and Female Hyperarticulated. On an individual trial, listeners heard a pair of sentences differing only in talker, and were asked to rate on a 1 – 7 scale the perceptual similarity of the two sentences. The sentence pairs always included talkers of the same gender producing the same sentence in the same style (e.g., two different male talkers producing citation speech or two different female talkers producing hyperarticulated speech). All possible talker combinations (within a gender), in both orders (i.e., A-B, B-A) were used, for a total of 168 trials. The interstimulus interval was 1 s. All trials were randomized for each participant and presented auditorily. Participants rated the sentence pairs by clicking on one of seven labeled buttons on a computer screen. No feedback was provided.

Results

The results of the voice similarity test, listed by stimulus set (Gender-Style) and individual sentence (Sentence 1 and Sentence 2), are given in Figure 5. For both male and female sets and for both sentences, voices in the citation style were judged as more similar to one another than voices in the hyperarticulated style. Of the four stimulus sets, the Female Citation voices were judged as more similar to one another than the Female Hyperarticulated voices. The difference in similarity ratings between the Male Citation and Male Hyperarticulated voices was smaller for Sentence 1 than Sentence 2. However, overall, Male Citation voices were still judged as more similar to one another than Male Hyperarticulated voices.

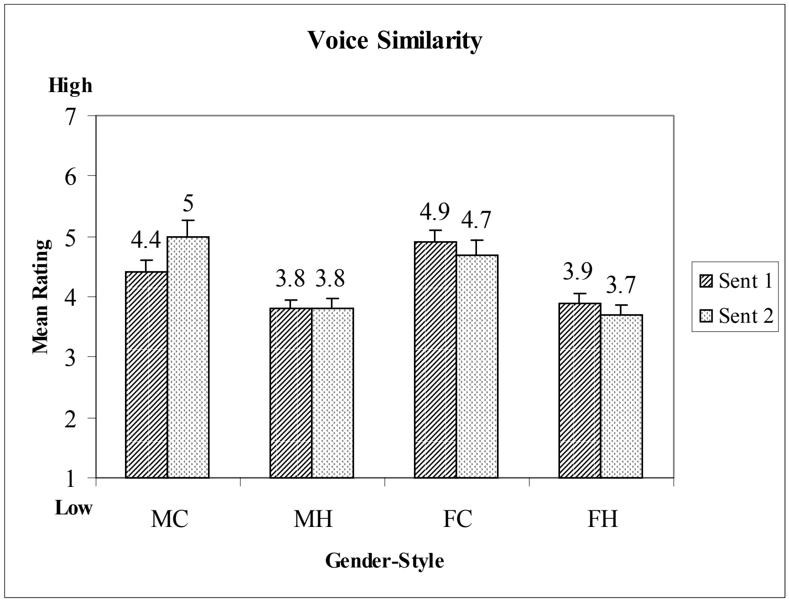

Figure 5. The mean similarity scores for the four stimulus sets, Male Citation (MC), Male Hyperarticulated (MH), Female Citation (FC), and Female Hyperarticulated (FH). Higher values denote greater similarity. "Sent" = Sentence. Error bars denote the standard error of the mean.

The mean perceptual ratings of the individual participants were submitted to a repeated-measures ANOVA, with two within-subjects factors (Gender, Speaking Style) and one between-subjects factor (Sentence). Of the main effects, only Speaking Style was significant (F (1, 30) = 64.9, $p < .0001$). Voices in a citation speaking style elicited significantly higher ratings than voices in a hyperarticulated speaking style. The Gender by Sentence (F (1, 30) = 16.2, $p < .0001$) and Gender by Speaking Style by Sentence (F (1, 30) = 11.3, $p < .01$) interactions were also significant.

The three-way interaction was further analyzed by running two two-way ANOVAs on the individual sentence data. For the participants who rated tokens of Sentence 1, both main effects and their interaction were significant (Gender: F (1, 15) = 13.4, $p < .01$; Speaking Style: F (1, 15) = 31.6, $p < .0001$; Gender by Speaking Style: F (1, 15) = 11.2, $p < .01$). In post hoc simple effects analyses, the effect of Speaking Style for both the male and female stimulus sets was significant ($p < 0.01$), while the effect of Gender within Speaking Style was only significant for the citation style stimuli ($p < 0.001$). For the participants who rated tokens of Sentence 2, Speaking Style was the only significant main effect. (F (1, 15) = 34, $p < .0001$), although Gender was marginally significant (F (1, 15) = 4.4, $p = .053$). The Gender by Speaking Style interaction was not significant.

In addition to analyzing the mean ratings, difference scores for the male and female voice sets were calculated to compare the effect of gender on the similarity of the citation and hyperarticulated speaking styles for each sentence. The mean rating for male and female hyperarticulated sentences of each subject was subtracted from the corresponding mean ratings for citation sentences (e.g., Listener 1's mean rating for Male Citation Sentence 1 minus that listener's mean rating for Male Hyperarticulated Sentence 1). The mean difference scores are shown in Figure 6. For Sentence 1, the mean difference score for the male stimulus set was lower than that for the female stimulus set, indicating that the effects of speaking style were larger for the female voices than the male voices in that sentence. In contrast, for Sentence 2, the mean difference score for the male stimulus set was higher than that for the female stimulus set.

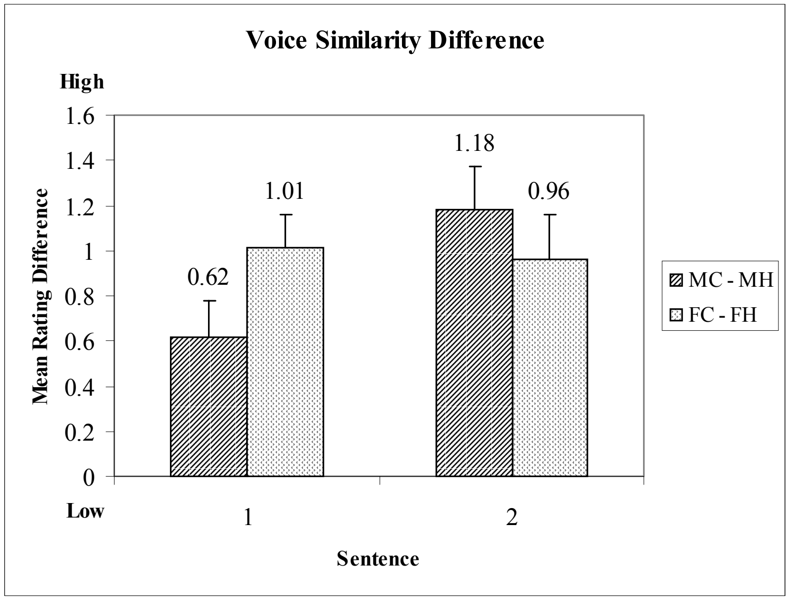

Figure 6. The mean difference scores for the two stimulus sets, Male Citation (MC) minus Male Hyperarticulated (MH) and Female Citation (FC) minus Female Hyperarticulated (FH). Higher values denote greater similarity. Error bars denote the standard error of the mean.

The difference scores of the individual participants were submitted to a repeated-measures ANOVA, with one within-subjects factor (Gender) and one between-subjects factor (Sentence). Neither of the main effects was significant. However, the Gender by Sentence interaction was significant (F (1, 30) = 11.3, $p < .01$). In post hoc simple effects analyses, the effect of Gender within both the male and female stimulus sets differed by sentence. For Sentence 1, the mean difference score for the male stimulus set was significantly lower than that for the female stimulus set ($p < 0.01$). For Sentence 2, the mean difference score for the male stimulus set was not significantly higher than that for the female stimulus set ($p = 0.09$).

General Discussion

The results of the perceptual similarity experiment using the scaling procedures were largely consistent with the acoustic properties of these speaking styles. Harnsberger and Goshert (2000) observed that sentences in citation and hyperarticulated speaking styles differed in overall duration, keyword duration, and vowel dispersion in the same manner for both male and female talkers. Likewise, listeners in the perceptual similarity experiment judged talkers producing citation speech as more similar to one another than talkers producing hyperarticulated speech, regardless of the gender of the talker. The greater similarity of the citation sentences suggests that voices in a citation speaking style should be harder to learn than the same voices in a hyperarticulated speaking style because they are perceptually less distinctive. We observed a significant effect of gender for Sentence 1 due to the relatively high mean similarity score elicited from female citation speech. However, this gender effect was not the same effect found in the first experiment (i.e., female hyperarticulated voices were easier to learn than female citation voices), because male citation and hyperarticulated speech also differed significantly in perceived similarity.

The gender effects observed in learning voices could be accounted for by the similarity difference scores shown in Figure 6. For one of the sentences, the difference between the female hyperarticulated and citation speaking styles was greater than the equivalent difference in the male speaking styles. It is possible that the female talkers produced a distinctive enough hyperarticulated speaking style that voice learning was influenced, while the male talkers made a more modest, though significant, distinction that was less informative. Given that the gender effect in the similarity difference scores varied by sentence, a scaling procedure for the entire stimulus set of Experiment 1 is required to determine whether the perceptual similarity of voices in Sentence 1 or 2 (or neither) were representative of the stimulus set as a whole. To further clarify the source of the gender effect, a replication of the first experiment is also needed with a new set of seven male and seven female talkers. In addition, other acoustic attributes of the stimuli should be examined, to ensure that male and female talkers manipulated the same gestural properties when shifting from citation to hyperarticulated speech. Such studies could take the form of a more extensive acoustic analysis, patterned after Brink et al. (1998) or Bradlow et al. (1996). Finally, a new voice learning experiment involving training over a longer period of time should be conducted, to determine if the gender effect on voice learning only emerges very early during stages of learning unfamiliar voices.

Regardless of the findings of subsequent studies, it is clear from the present results that speaking style influences the learning of novel voices. The results of the present study show that the relationship between speaking style and voice learning is complex, and may involve both attributes of the signal as well as listeners' prior experience with different kinds of voices. Further studies would serve to clarify our understanding of this process and, in turn, enrich our understanding of the nature of the perceptual learning process.

Summary and Conclusions

This study examined the effects of two speaking styles on the learning of novel voices. Listeners participated in a voice learning experiment consisting of four phases, (1) training, (2) test, (3) sentence generalization, and (4) speaking style generalization. In all four phases, listeners were presented with sentences produced by either seven male or seven female talkers in either a citation or hyperarticulated speaking style. The experiment was designed to assess two hypotheses concerning the effects of speaking style on novel voice learning: the *clear speech hypothesis* and the *idiosyncratic articulation hypothesis*. According to the *clear speech hypothesis*, hyperarticulated voices should be easier to learn than citation voices because hyperarticulated speech has been shown in prior work to be highly informative of other aspects of the speech signal, particularly, its linguistic content. According to the *idiosyncratic articulation hypothesis*, citation voices should be easier to learn than hyperarticulated voices because idiosyncratic gestural styles emerge that are normally suppressed in hyperarticulated speech. The results supported neither of these hypotheses directly because of an unanticipated effect of gender on the voice identification results. As it happened, female voices were easier to learn in a hyperarticulated style relative to a citation style in the training and test phases. In the style generalization phase with female voices, training with the hyperarticulated tokens provided a greater advantage in identifying voices in a novel style relative to training with citation tokens. In contrast, no differences were observed in the learning of male voices in different styles.

Several accounts of the gender effect were offered, including the possible role of prior experience in listening to male and female voices, which have been reported to differ in terms of the frequency of occurrence of casual and hyperarticulated speech forms. The gender interaction in voice learning was examined further in a similarity scaling experiment in which listeners rated the similarity of male and female voices in citation and hyperarticulated speaking styles. Both male and female citation sentences were rated as significantly more similar than the corresponding hyperarticulated sentences. However, for one sentence, the difference in mean similarity for the female citation and hyperarticulated voices was significantly greater than the corresponding difference in the male voices, indicating that the female talkers may have produced a perceptually more distinct hyperarticulated style than the male talkers. Such a distinct style may have contributed to the gender effect observed in the voice learning experiment. Overall, the results show that speaking style affects voice learning, although the interpretation of the results is complicated by differences in learning male and female novel voices in the laboratory using this experimental paradigm.

References

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

Bricker, P.D. & Pruzansky, S. (1976). Speaker recognition. In N.J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 295-326). New York: Academic Press.

Brink, J., Wright, R., & Pisoni, D.B. (1998). Eliciting speech reduction in the laboratory: Assessment of a new experimental method. In *Research on Spoken Language Processing Progress Report No. 22* (pp. 396-420). Bloomington, IN: Speech Research Laboratory, Indiana University.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, *15*, 39-54.

Fischer, J.L. (1958). Social influences on the choice of a linguistic variant. *Word*, *14*, 47-56.

Goggin, J.P, Thompson, C.P., Strube, G., & Simental, L.R. (1991). The role of language familiarity in voice identification. *Memory and Cognition, 19*, 448-458.

Harnsberger, J.D. & Goshert, L. (2000). Reduced, citation, and hyperarticulated speech in the laboratory: An acoustic analyses. In *Research on Spoken Language Processing Progress Report No. 24* (pp. 357-368). Bloomington, IN: Speech Research Laboratory, Indiana University.

Harnsberger, J.D. & Pisoni, D.B. (1999). Eliciting speech reduction in the laboratory II: Calibrating cognitive loads for individual talkers. In *Research on Spoken Language Processing Progress Report No. 23* (pp. 339-349). Bloomington, IN: Speech Research Laboratory, Indiana University.

Hollien, H., Majewski, W., & Doherty, E.T. (1982). Perceptual identification of voices under normal, stress, and disguise conditions. *Journal of Phonetics*, *10*, 139-148.

Johnson, K., Ladefoged, P., & Lindau, M. (1993). Individual differences in vowel production. *Journal of the Acoustical Society of America, 94,* 701-714.

Kalikow, D.N., Stevens, K.N., & Elliot, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61,* 1337-1351.

Kreiman, J. (1997). Listening to voices: Theory and practice in voice perception research. In K. Johnson and J.W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 85-108). San Diego: Academic Press.

Kreiman, J., Gerratt, B.R., Precoda, K. & Berke, G.S. (1992). Individual differences in voice quality perception. *Journal of Speech and Hearing Research*, *35*, 512-520.

Kuwabara, H. & Takagi, T. (1991). Acoustic parameters of voice individuality and voice-quality control by analysis-synthesis method. *Speech Communication*, *10*, 491-495.

Labov, W. (1970). The study of language in its social context. *Studium Generale*, *23*, 30-87.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H & H theory. In W.J. Hardcastle and A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403-439). Dordrecht: Kluwer Academic Publishers.

Mattingly, I. (1966). Speaker variation and vocal-tract size. *Journal of the Acoustical Society of America, 39,* 1219.

Moon, S.-J. & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America, 96,* 40-55.

Murray, T. & Singh, S. (1980). Multidimensional analysis of male and female voices. *Journal of the Acoustical Society of America, 68,* 1294-1300.

Nygaard, L.C. & Pisoni, D.B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics*, *60*, 355-376.

Picheny, M.A., Durlach, N.I., & Braida, L.D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, *28*, 96-103.

Picheny, M.A., Durlach, N.I., & Braida, L.D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, *29*, 434-446.

Singh, S. & Murray, T. (1978). Multidimensional classification of normal voice qualities. *Journal of the Acoustical Society of America, 64,* 81-87.

Tartter, V.C. (1991). Identifiability of vowels and speakers from whispered syllables. *Perception and Psychophysics*, *49*, 365-372.

Trudgill, P. (1974). *The social differentiation of English in Norwich. Cambridge*: Cambridge University Press.

Thompson, C.P. (1987). A language effect in voice identification. *Applied Cognitive Psychology*, *1*, 121-131.

Van Lancker, D., Kreiman, J., & Wickens, T.D. (1985). Familiar voice recognition: Patterns and parameters. Part II: Recognition of rate-altered voices. *Journal of Phonetics*, *13*, 39-52.

Appendix A:

The SPIN sentences recorded for this study

The farmer harvested his crop.
His boss made him work like a slave.
He caught the fish in his net.
Close the window to stop the draft.
The beer drinkers raised their mugs.
I made the phone call from a booth.
The cut on his knee formed a scab.
The railroad train ran off the track.
They drank a whole bottle of gin.
The airplane dropped a bomb.
I gave her a kiss and a hug.
The soup was served in a bowl.
The cookies were kept in a jar.
How did your car get that dent?
The baby slept in his crib.
The cop wore a bulletproof vest.
No one was injured in the crash.
The hockey player scored a goal.
How long can you hold your breath?
At breakfast he drank some juice.
The king wore a golden crown.
He got drunk in the local bar.
The doctor prescribed the drug.
The landlord raised the rent.
Playing checkers can be fun.
Throw out all this useless junk.
Her entry should win first prize.
The stale bread was covered with mold.
I ate a piece of chocolate fudge.
The story had a clever plot.
He's employed by a large firm.
The mouse was caught in the trap.
I've got a cold and a sore throat.
The judge is sitting on the bench.

Appendix B:

The SPIN sentences presented in the first experiment

Sentences used in the Training and Test Phases

- I made the phone call from a booth.
- The railroad train ran off the track.
- No one was injured in the crash.
- The landlord raised the rent.
- The beer drinkers raised their mugs.

Sentences used in the Sentence Generalization Phase

- Playing checkers can be fun.
- Her entry should win first prize.
- The stale bread was covered with mold.
- He's employed by a large firm.
- The judge is sitting on the bench.

Sentences used in the Speaking Style Generalization Phase

- His boss made him work like a slave.
- They drank a whole bottle of gin.
- The hockey player scored a goal.
- How long can you hold your breath?
- The doctor prescribed the drug.