

NIH Public Access

Author Manuscript

J Speech Lang Hear Res. Author manuscript; available in PMC 2013 February 24.

Published in final edited form as:

J Speech Lang Hear Res. 2011 August ; 54(4): 1240–1245. doi:10.1044/1092-4388(2010/10-0168).

Perceptual Adaptation of Voice Gender Discrimination with Spectrally Shifted Vowels

Tianhao Li and **Qian-Jie Fu**

House Research Institute, 2100 West Third Street, Los Angeles, CA

Abstract

Purpose—To determine whether perceptual adaptation improves voice gender discrimination of spectrally shifted vowels and, if so, which acoustic cues contribute to the improvement.

Method—Voice gender discrimination was measured for 10 normal-hearing subjects, during 5 days of adaptation to spectrally shifted vowels, produced by processing the speech of 5 male and 5 female talkers with 16-channel sine-wave vocoders. The subjects were randomly divided into 2 groups; one subjected to 50-Hz, and the other to 200-Hz, temporal envelope cutoff frequencies. No preview or feedback was provided. Results: There was significant adaptation in voice gender discrimination with the 200-Hz cutoff frequency, but significant improvement was observed only for 3 female talkers with F0 > 180 Hz and 3 male talkers with F0 < 170 Hz. There was no significant adaptation with the 50-Hz cutoff frequency.

Conclusions—Temporal envelope cues are important for voice gender discrimination under spectral shift conditions with perceptual adaptation, but spectral shift may limit the exclusive use of spectral information and/or the use of formant structure on voice gender discrimination. The results have implications for cochlear implant users and for understanding voice gender discrimination.

Keywords

perceptual adaptation; voice gender discrimination; spectrally shifted speech

Introduction

Cochlear implants (CIs) have been relatively successful as auditory prosthetic devices, allowing people with sensorineural deafness to recover partial hearing. CI users achieve a high level of speech recognition performance in ideal listening environments but have difficulties with pitch-related listening tasks (e.g., music melody recognition, speaker/voice identification, tone recognition). CI devices in their current configuration generally provide users with spectrally shifted and/or distorted signals, with limited spectral and temporal resolution. Due to reduced spectral resolution, harmonic complex tones are not available for the perception of pitch. CI users rely mainly on tonotopic information and temporal envelope cues for pitch-related perception (Moore & Carlyon, 2005).

Previous CI simulation studies have investigated the relative contribution of spectral information and temporal envelope cues to pitch-related perception (Fu, Chinchilla, & Galvin, 2004; Kong, Cruz, Jones, & Zeng, 2004; Kong & Zeng, 2006; Xu, Tsai, & Pfingst, 2002). However, most of these studies used tonotopically matched CI simulations. In the real world, CI users usually receive tonotopically shifted and distorted signals, due to

Corresponding Author: Tianhao Li, House Research Institute, 2100 West Third Street, Los Angeles, CA, tianhaol@gmail.com.

device-or patient-related factors. Forsome CI users, however, pitch perception based on tonotopic information can shift as much as two octaves over a period of several years (Reiss, Turner, Erenberg, & Gantz, 2007). Therefore, it is necessary not only to understand the relative importance of spectral information and temporal envelope cues to pitch-related perception under tonotopic distortion but also to investigate the effects of perceptual adaptation.

The present study examined perceptual adaptation of voice gender discrimination with spectrally shifted vowels. Voice gender discrimination is a pitch-related perception task. Previous studies of normal-hearing (NH) listeners showed that fundamental frequency (F0) and formant structure are major factors for voice gender discrimination (Bachorowski & Owren, 1999; Fellowes, Remez, & Rubin, 1997) and that F0 is somewhat more important than formant structure (Hillenbrand & Clark, 2009). Similarly, previous studies using acoustic CI simulations found that both temporal envelope cues and spectral cues supported voice gender discrimination. However, their relative importance was affected by spectral resolution, spectral mismatch, and the range of intertalker F0s (Fu et al., 2004; Fu, Chinchilla, Nogaki, & Galvin, 2005; Gonzalez & Oliver, 2005). Accordingto Fu et al. (2005), temporal envelope cues play an important role in voice gender discrimination when the F0s of male and female talkers are widely separated, especially with reduced spectral resolution or spectral mismatch; however, spectral informationbecomes more important when the range of intertalker F0s is small. Furthermore, for a small intergender F0 difference, voice gender discrimination performance declines after an octave upward spectral shift, even with temporal envelope cues. To evaluate the relative contribution of temporal envelope and spectral cues to voice gender discrimination under tonotopic distortion after perceptual adaptation, we used stimuli that were generated from a set of vowels with a small intergender F0 difference and that varied in the amount of spectral and/ or temporal information. The present study addressed two specific questions: (a) Can shortterm perceptual adaptation improve pitch-related perception under spectral mismatch? (b) Do temporal envelope cues or tonotopic information contribute more strongly to the improvement?

Method

Subjects

Ten NH subjects (5 males and 5 females, 18–33 years old) participated in the study; all were native speakers of American English. All subjects had pure-tone thresholds better than 20 dB HL at octave frequencies of 125–8000 Hz. None of the subjects had prior experience with acoustic CI simulations. The Institutional Review Board of St. Vincent Medical Center approved the involvement of human subjects in the present project. All subjectsgave informed consent.

Speech Materials

Twelve medial vowels, presented in the format of /h-V-d/ (had, head, hod, hawed, hayed, heard, hid, heed, hud, hood, who'd, and hoed), were used as speech stimuli. Five male and five female talkers were selected from speech samples recorded by Hillenbrand, Getty, Clark, and Wheeler (1995) and were similar to those in Talker Set 2 of Fu et al. (2005). The F0s of the five male (M1 M5) and five female (F1 F5) talkers overlapped within the range of 150–200 Hz (see Table 1). The mean F0 values of male and female talkers differed by 10 Hz, which is small relative to the mean difference between F0s of natural male and female talkers (90 Hz).

Signal Processing and Test Conditions

We used 16-channel sine-wave vocodersto generate speech stimuli, in order to reduce floor effects of limited spectral resolution on voice gender discrimination. Voice gender discrimination by CI users was more consistent with acoustic CI simulations produced by sine-wave vocoders than by noise-band vocoders (Fu et al., 2004), and 16 spectral channels were required for good voice gender discrimination with insufficient temporal envelope cues (Fu et al., 2005).

The speech signal input into the 16-channel sine-wave vocoder had an acoustic frequency range of 200–7000 Hz and was band-pass filtered into 16 channels after pre-emphasis. Table 2 shows the corner and center frequencies of the 16 analysis filters (fourth-order Butterworth filters). The spatial distribution of center frequencies was calculated according to Greenwood's (1990) function, assuming a 35-mm-long cochlea. In each channel, the temporal envelope was extracted by half-wave rectification and a low-pass, fourth-order Butterworth filter. The cutoff frequency of the low-pass filters was50 or 200 Hz, removing or preserving the higher rate temporal envelope cues, respectively. The temporal envelopes in each channel were then used to modulate sine-wave carriers that were either tonotopically matched or shifted relative to the analysis bands (see Table 2). For the shifted condition, the frequency range of the overall carrier bands was 683–7408 Hz, simulating a 16-mm implanted electrode array. The lowest frequency channel was upwardly shifted by 6.2 mm, close to the moderate shift condition (6 mm) tested by Li, Galvin, and Fu (2009), which showed that vowel recognition could be improved for this shift condition with eight spectral channels and without explicit training. Finally, the modulated carriers were summed and normalized to have the same long-term root-mean-square as the input speech signal.

Procedure

The 10 subjects were randomly divided into two groups: Group 1 (four subjects) learned voice gender discrimination with spectrally shifted vowel tokens without the higher rate temporal envelope cues (50-Hz cutoff frequency); Group 2 (six subjects) learned voice gender discrimination with spectrally shifted vowel tokens with temporal envelope cues (200-Hz cutoff frequency). Subjects were seated in a sound-treated booth and listened to the speech stimuli via a loudspeaker. The presentation level of all speech stimuli was 65 dB SPL. Voice gender discrimination was measured using a two-alternative forced choice paradigm. A vowel token was randomly selected from the stimulus set (12 Vowels x 10 Talkers = 120 Vowel Tokens) and presented to subjects, with no repetition. Following the presentation of each test token, subjects responded by clicking one of two response buttons, labeled "male" and "female."No preview or feedback was provided. There were 120 tokens in each trial, and performance was quantified as the percentage of correct responses.

To lessen the effect of the top-down process, the adaptation protocol used an unsupervised, 5-day learning paradigm, similar to the protocol used by Li et al. (2009). Pre-adaptation baseline measures of voice gender discrimination were obtained on Day 1, for both spectrally matched and shifted vowel tokens, with temporal envelope cues and without temporal envelopecues. During the pre-adaptation measurement session, voice gender discrimination was measured for spectrally matched vowels prior to spectrally shifted vowels. Performance was remeasured for all baseline conditions immediately after the 5-day adaptation phase. During the postadaptation measurement session, voice gender discrimination was measured for spectrally shifted vowels prior to spectrally matched vowels. To allow subjects in Group 1 to adapt to spectrally shifted vowels without temporal envelope cues (50-Hz cutoff frequency), we tested the spectrally shifted vowels with the 200-Hz cutoff frequency first, followed by the spectrally shifted vowels with the 50-Hz cutoff frequency. For subjects in Group 2, the test order was reversed. During the 5-day

adaptation period, Group 1 was tested daily using only spectrally shifted vowels with the 50- Hz cutoff frequency, and Group 2 was tested daily using only spectrally shifted vowels with the 200-Hz cutoff frequency. Five trials were administered daily. No preview and no feedback were provided during the adaptation period.

Results

Subjects were able to partially adapt to spectrally shifted speech with temporal cues up to 200 Hz for voice gender discrimination (Group 2) but not to the shifted speech without the higher rate temporal envelope cues (Group 1). Mean performance for Group 1 on Days 1–5 ranged from 58% to 62% (see Figure 1) and showed no significant improvement in voice gender discrimination during the adaptation period: one-way repeated measures analysis of variance (ANOVA), $F(4, 12) = 1.815$, $p = .191$. Consistent with the overall performance, there was no significant improvement in voice gender discrimination for any individual talker during the adaptation period across the four subjects in Group 1 (see Table 1). In contrast, the mean performance of Group 2 gradually increased from 60% correct on Day 1 to 73% correct on Day 5 (see Figure 1), showing significant improvement during the adaptation period: one-way repeated measures ANOVA, $F(4, 20) = 9.115$, p < .001. The significant change occurred by Day 3, after which there was no further improvement. Despite the overall improvement, one-way repeated measures ANOVAs showed that voice gender discrimination performance significantly improved for only six out of 10 talkers (see Table 1). On Day 5, performance was positively correlated with the F0 value for the five female talkers ($r = .963$, $p = .008$) and negatively correlated with the F0 value for the five male talkers ($r = -.886$, $p = .045$).

We used t tests to compare performance between groups under specific test conditions; we used paired t tests to compare pre-and postadaptation performance within the group and performance within the group under different spectral shift conditions. Significance of differences was accepted at $p < .05$. The pre-adaptation baseline performances of Groups 1 and 2 were not significantly different (see Figure 2) for all conditions. Group 1 showed no significant change between pre-and postadaptation performance for all test conditions. On Day 5, performance for the spectrally shifted condition with temporal envelope cues was still significantly poorer than performance for two spectrally matched conditions, and there was no significant difference between the two spectrally matched conditions. Group 2 showed improvement in postadaptation performance for spectrally shifted speech, bothwith and without temporal envelope cues. However, the mean pre-adaptation performance of Group 2 for spectrally shifted speech without temporal envelope cues was 6% lower, and postadaptation performance was only 1% higher than for Group 1. Thus, the significant improvement observed under this condition might be attributable to the pre-adaptation measurements being made prior to the spectrally shifted condition with a 200-Hz cutoff frequency and/or to intersubject variability within a small group of subjects. There was no significant change in performance after adaptation for the spectrally matched condition with temporal envelope cues. However, when temporal envelope cues were removed, postadaptation performance under the spectrally matched condition declined significantly compared with pre-adaptation performance. It is possible that the improved performance of Group 2 during the adaptation process came from the use of shifted temporal envelope cues and that adapted subjects had difficulty using only spectral information, especially when measured after the spectrally shifted conditions. After the 5-day adaptation period, performance for the spectrally shifted condition with temporal envelope cues was not significantly different than for the spectrally matchedcondition with or without temporal envelope cues.

Discussion

Results ofthe present study showed that listeners were able to adapt to spectrally shifted speech for voice gender discrimination only when temporal envelope cues were present; the exclusive use of spectral cues for voice gender perception was limited by spectral mismatch, even after unsupervised, short-term perceptual adaptation. Moreover, the postadaptation performance of subjects in Group 2 for spectrally shifted conditions with temporal envelope cues was positively correlated with F0 for five female talkers and negatively correlated with F0 for five male talkers. This suggests that after the 5-day adaptation period, temporal envelope periodicity cues played a primary role in discriminating voice gender of spectrally shifted vowels. Temporal envelope cues are important to voice gender discrimination when there is a large difference in F0s between male and female talkers, especially with limited spectral resolution or with spectral mismatch (Fu et al., 2005). Results presented herein demonstrated that with perceptual adaptation, temporal envelope cues are also important when there is tonotopic mismatch or a small range of F0s, even though the importance of temporal envelope cues is lessened under these conditions (Fu et al., 2005). These results also suggest that with sufficient perceptual adaptation, temporal envelope cues are somewhat independent of tonotopic information for voice gender discrimination or pitchrelated perception.

Although subjects showed the ability to adapt to spectrally shifted speech with a 200-Hz cutoff frequency without explicit training, it was unclear whether subjects made use of temporal envelope cues with temporal mechanisms or spectral mechanisms. As discussed by Gonzalez and Oliver (2005), gender discrimination performance is better withsine-wave vocoders than with noise-band vocoders, partly because amplitude modulation detection is easier with sine-wave carriers than with noise carriers (Kohlrausch, Fassel, & Dau, 2000; Viemeister, 1979) and partly because sine-wave vocoders introduce resolved side bands. The side bands centered at sine-wave carriers may provide F0 information— that is, spectral mechanisms for voice gender discrimination. Although sine-wave vocoders were used in the present study, it is likely that subjects mainly used temporal mechanisms to improve performance. The first five channels of the 16-channel sine-wave vocoder were comparable to the F0 range of the talkers, which may have inhibited the side bands needed for spectral mechanisms. A recent study also suggested that sine-wave vocoders with high temporal cutoff frequencies provided more periodicity and intonation information than did noise-band vocoders (Souza & Rosen, 2009).

Subjects in the present study did not adapt to spectrally shifted speech for voice gender discrimination. These results and results of previous studies showed that without spectral mismatch, 10-to 16-channel spectral bands alone provide sufficient information for good voice gender discrimination (Fu et al., 2005; Gonzalez & Oliver, 2005). However, voice gender discrimination dramatically dropped after spectral shifting, even after short-term unsupervised perceptual adaptation. Thus, the exclusive use of spectral information for voice gender identification is limited by spectral shift, at least with reduced spectral resolution and/or with short-term perceptual adaptation, suggesting that some factors related to tonotopic cues affect voice gender discrimination and perceptual adaptation. It is possible that subjects identified spectrally shifted speech with an upward shift of spectral profiles as female. However, confusion matrix analyses displayed no obvious shift in "female" response. Alternatively, spectral profiles may support voice gender discrimination by providing perception of sound quality or formant structure. Spectral shift may affect perception of the sound quality or the formant structure, thereby affecting voice gender discrimination. For pitch-related perception, furthermore, the rate of adaptation to changes in sound quality or formant structure may be slower than the rate of adaptation to shifted temporal envelope cues. Thus, there was no significant adaptation to shifted spectral

information within the 5-day adaptation period without temporal envelope cues. In contrast, Reiss et al. (2007) found that tonotopic information-based pitch perception could shift up to two octaves over years, possibly due to long-term, higher-level changes. More feedback may facilitate adaptation to changes of sound quality or formant structure.

Results of the present study provide insights into voice gender discrimination by human listeners and have implications for CI users. While F0 has already been shown to be somewhat more important than formant structure (Hillenbrand & Clark, 2009), results presented herein show that voice gender discrimination can use F0 independently of tonotopic information after short-term unsupervised adaptation. However, the results also indicate that the relative roles of F0 and formant structure can be affected by the F0 of the talkers, as voice gender discrimination by Group 2 improved only for male talkers with F0 < 170 Hz and for female talkers with F0 > 180 Hz. The 170-to 180-Hz range is the median F0 of natural male and female talkers and appears to be the dividing line forF0-based voice gender discrimination. Formant structure might be more important for voice gender discrimination for a female talker with $F0 < 180$ Hz and a male talker with $F0 > 170$ Hz. However, the use of formant structure in voice gender discrimination may be limited by spectral shift and/or limited spectral resolution. For CI users, who have difficulties in pitchrelated listening tasks and in speech recognition in background noise, results of the present study indicate that spectral mismatch might be another factor, in addition to the limited spectral and temporal resolution, affecting performance of CI users in complex, pitch-related perception tasks. Thus, more independent channels, more precise frequency-to-place mapping, or explicit training in the exclusive use of tonotopic information will be helpful for CI users faced with complex, pitch-related perception tasks.

Acknowledgments

This research was supported by National Institute on Deafness and Other Communication Disorders Grant DC004792. We wouldlike to thank all subjects for their time and attention.

References

- Bachorowski JA, Owren M. Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. The Journal of the Acoustical Society of America. 1999; 106:1054–1063. [PubMed: 10462810]
- Fellowes JM, Remez RE, Rubin PE. Perceiving the sex and identity of a talker without natural vocal timbre. Perception & Psychophysics. 1997; 59:839–849. [PubMed: 9270359]
- Fu QJ, Chinchilla S, Galvin JJ. Voice gender discrimination and vowel recognition in normal-hearing and cochlear implant users. Journal of the Association for Research in Otolaryngology. 2004; 5:253–260. [PubMed: 15492884]
- Fu QJ, Chinchilla S, Nogaki G, Galvin JJ. Voice gender identification by cochlear implant users: The role of spectral and temporal resolution. The Journal of the Acoustical Society of America. 2005; 118:1711–1718. [PubMed: 16240829]
- Gonzalez J, Oliver JC. Gender and speaker identification as a function of the number of channels in spectrally reduced speech. The Journal of the Acoustical Society of America. 2005; 118:461–470. [PubMed: 16119365]
- Greenwood DD. A cochlear frequency-position function for several species— 29 years later. The Journal of the Acoustical Society of America. 1990; 87:2592–2605. [PubMed: 2373794]
- Hillenbrand JM, Clark MJ. The role of f(0) and formant frequencies in distinguishing the voices of men and women. Attention, Perception & Psychophysics. 2009; 71:1150–1166.
- Hillenbrand JM, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. The Journal of the Acoustical Society of America. 1995; 97:3099–3111. [PubMed: 7759650]

- Kohlrausch A, Fassel R, Dau T. The influence of carrier level and frequency on modulation and beatdetection thresholds for sinusoidal carriers. The Journal of the Acoustical Society of America. 2000; 108:723–734. [PubMed: 10955639]
- Kong YY, Cruz R, Jones JA, Zeng FG. Music perception with temporal cues in acoustic and electric hearing. Ear and Hearing. 2004; 25:173–184. [PubMed: 15064662]
- Kong YY, Zeng FG. Temporal and spectral cues in Mandarin tone recognition. The Journal of the Acoustical Society of America. 2006; 120:2830–2840. [PubMed: 17139741]
- Li T, Galvin JJ, Fu QJ. Interactions between unsupervised learning and the degree of spectral mismatch on short-term perceptual adaptation to spectrally shifted speech. Ear and Hearing. 2009; 30:238–249. [PubMed: 19194293]
- Moore, BCJ.; Carlyon, RP. Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In: Plack, CJ.; Oxenham, AJ.; Fay, RR., editors. Pitch: Neural coding and perception. New York, NY: Springer; 2005. p. 234-277.
- Reiss LAJ, Turner CW, Erenberg SR, Gantz BJ. Changes inpitch with a cochlear implant over time. Journal of the Association for Research in Otolaryngology. 2007; 8:241–257. [PubMed: 17347777]
- Souza P, Rosen S. Effects of envelope bandwidth on the intelligibility of sine-and noise-vocoded speech. The Journal of the Acoustical Society of America. 2009; 126:792–805. [PubMed: 19640044]
- Viemeister NF. Temporal modulation transfer functions based upon modulation thresholds. The Journal of the Acoustical Society of America. 1979; 66:1363–1380.
- Xu L, Tsai Y, Pfingst BE. Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. The Journal of the Acoustical Society of America. 2002; 112:247–258. [PubMed: 12141350]

Li and Fu Page 8

Figure 1.

Voice gender discrimination performance $(M \pm SE)$ for spectrally shifted vowels with 50-Hz or 200-Hz temporal envelope cutoff frequencies (freq), as a function of test day. Random chance of correct performance is 50%.

Figure 2.

Pre-and postadaptation baseline voice gender discrimination performance $(M + SE)$ for two groups. Random chance of correct performance is 50%. Error bars represent 1 SE.

NIH-PA Author Manuscript

Table 1

Mean voice gender discrimination performance $(\pm$ SE) for each talker on Days 1-5, averaged across subjects in Groups 1 and 2. Mean voice gender discrimination performance (± SE) for each talker on Days 1–5, averaged across subjects in Groups 1 and 2.

fundamental frequency.

fundamental frequency.

 NIH-PA Author ManuscriptNIH-PA Author Manuscript **Table 2**

