Published in final edited form as:

Ear Hear. 2012 May ; 33(3): 411–420. doi:10.1097/AUD.0b013e31823d78dc.

Gender identification in younger and older adults: use of spectral and temporal cues in noise-vocoded speech¹

Kara C. Schwartz, Au.D., Ph.D. and Monita Chatterjee, Ph.D.

Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland, 20742, USA

Abstract

Objective—The aim of the current study was to investigate potential effects of age on the ability of normal-hearing (NH) adult listeners to utilize spectral and temporal cues when performing a voice gender identification task.

Design—Ten younger and ten older NH adult listeners were measured on their ability to correctly identify the speaker gender of six different vowel tokens (H-/vowel/-D) when spoken by eight speakers (four male, four female). Spectral (number of channels) and temporal cues (low-pass cut-off frequency for temporal envelope extraction) were systematically manipulated using noiseband vocoding techniques; stimuli contained 1, 4, 8, 16 or 32 spectral channels, while the low-pass cutoff frequency of the temporal envelope filter was 20, 50, 100, 200 or 400 Hz. Further, the fundamental frequencies (F0s) of the vowel tokens were manipulated to create two conditions: ‘Expanded’ (large range of F0 values), and ‘Compressed’ (small range of F0 values).

Results—In general, younger listeners performed better than the older listeners, but only when stimuli were spectrally degraded. For both the Expanded and Compressed conditions, the overall performance of the younger listeners was better than that of the older listeners, suggesting age-related deficits in both spectral and temporal processing. Furthermore, a significant interaction between age group and temporal envelope cues revealed that older listeners received less benefit from increasing temporal envelope information compared to the benefit observed among younger listeners. Specifically, the performance of the younger NH group (collapsed across number of channels), but not the older NH group, improved as the temporal envelope cut-off frequency was increased from 50 to 400 Hz.

Conclusions—The results reported here support previous findings of senescent declines in perceiving spectrally reduced speech and temporal amplitude modulation processing. These results suggest that when F0 values are similar to one another, younger listeners can use temporal cues alone to glean voice-pitch information but older listeners exhibit a lessened ability to use such cues. Previous studies have demonstrated the importance of temporal envelope cues in periodicity perception (e.g, gender recognition) by cochlear-implant (CI) listeners. The results of the present study suggest that that aging affects the use of such cues, and consequently gender recognition might be poorer among older CI recipients.

¹Portions of this work were presented at The 2009 Meeting of the American Auditory Society, Scottsdale, AZ.

Author to whom correspondence should be addressed: Kara Schwartz, University of Maryland, 0100 LeFrak Hall, College Park, MD 20742. **Address to where reprints should be sent:** Monita Chatterjee, University of Maryland, 0100 LeFrak Hall, College Park, MD 20742.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

aging; cochlear implants; pitch

Introduction

Voice-gender identification and discrimination play a significant role in everyday communication and multi-talker listening environments. Normal-hearing (NH) individuals are able to utilize voice-pitch cues to separate sound sources and extract a target talker from other distracting voices (Arehart et al., 1997; Summers & Leek, 1998; Brungart, 2001). Although multiple factors contribute to the perception of voice gender, the fundamental frequency (F0) of the talker commonly prevails as the most important cue (Murry & Singh, 1980; Klatt & Klatt, 1990). Researchers have debated the importance of spectral and temporal cues to pitch perception (Goldstein, 1973; Licklider, 1951; Meddis & Hewitt, 1991a; 1991b; Schouten, 1940; Terhardt, 1972; von Hemholtz, 1877; Wightman, 1973). While it is beyond the scope of this paper to review such theories, it has been repeatedly demonstrated that pitch perception is strongest when the lower harmonics are resolved (Ritsma, 1967; Shackleton & Carlyon, 1994; Qin & Oxenham, 2005); this ability, of course, requires adequate spectral resolution. However, these studies also demonstrate that the pitch of a harmonic complex can also be roughly perceived when the lower harmonics are removed (or unresolved) by utilizing periodicity information conveyed in the temporal envelope.

It has been repeatedly demonstrated that those with reduced spectral resolution, such as cochlear-implant (CI) users, demonstrate poor gender discrimination abilities (Fu et al., 2004; 2005; Gonzalez & Oliver, 2005; Vongphoe & Zeng, 2005). This is presumably due to the absence of temporal fine structure information and broadened auditory filters in CI users (primarily attributed to electrical current overlap). It is also important to note that information about talker-gender is available using formant-structure cues (Childers & Wu, 1991) via spectral peaks. Studies have shown, however, that spectral-peak resolution is poor among CI users (Henry & Turner, 2003; Henry, Turner, & Behrens, 2005). Recent evidence obtained among children suggests that, while children with CIs can perceive differences in voice pitch, spectral envelope cues do not strongly promote differentiation of talker-gender in CI users within this age group (Kovacic & Balaban, 2009; Vongpaisal et al., 2010). In fact, overall evidence suggests that CI users rely primarily on temporal envelope cues to discern voice pitch (Chatterjee & Peng, 2008; Fu et al., 2004, 2005; Kovacic & Balaban, 2009).

Although temporal envelope periodicity cues result in weak F0 perception, several studies have demonstrated that CI users are able to discern differences in talker gender (primarily based on differences in F0 information). In such studies, performance was measured among CI listeners or among NH listeners using CI simulation techniques such as noiseband or sinewave vocoding. These methods mimic CI processing by systematically limiting the number of spectral channels and temporal envelope information available to the listener. Briefly, the original signal is divided into a given number of frequency bands then replaced with noisebands or sinewaves comprised of frequencies that approximate the original bandwidth. Each frequency band is then modulated with its respective temporal envelope, which can be manipulated (i.e., low-pass filtered) or left unaltered. The frequency bands are then summed to create the final vocoded signal (Shannon et al., 1995). Using this method, the number of channels (spectral cue) and low-pass cut-off frequency for temporal envelope extraction (temporal cue) can be systematically manipulated.

For example, Fu et al. (2004) measured gender identification in six younger NH adults (ages 22-30) when listening to sinewave vocoded speech and in 11 adult CI users (ages 39-70). Voice gender identification was measured using a closed-set, two-alternative forced-choice (2AFC) task. Stimuli consisted of 12 vowels varying in talker gender. For NH participants, the degree of spectral resolution was varied by providing 4-32 spectral channels, and the upper-limit of the temporal envelope was manipulated as well (20-320 Hz), thereby allowing access to varying degrees of temporal envelope cues. Results showed that performance was near perfect with 32 channels, but listeners exhibited systematic declines in performance as the number of channels was reduced. Further, performance improved as the temporal envelope cut-off frequency was increased (from 20 to 320 Hz) particularly when spectral cues were severely degraded thereby exhibiting a significant interaction between spectral and temporal cues. Results from subsequent, related studies corroborated those obtained by Fu et al., (2004) (Fu et al., 2005; Gonzolez & Olivier, 2005), and Fu et al. (2005) further demonstrated that the ability to utilize residual temporal envelope cues is reduced when F0 values are more similar across talkers. These results agreed with those obtained by Vongphoe & Zeng (2005) who found that CI listeners (and NH listening to CI simulations) experienced difficulty utilizing temporal envelope cues when required to discriminate between small differences in F0 in performing a talker identification task.

One important detail regarding the above studies (Fu et al., 2004; 2005; Vongphoe & Zeng, 2005) is the fact that a *sinewave* vocoding method was employed (instead of a *noiseband* vocoding method) the former of which often results in better performance on several tasks (Dorman et al., 1997; Gonzalez & Oliver, 2005) particularly when task performance was dependent upon F0 coding. Gonzalez & Oliver (2005) measured gender identification in 15 NH listeners (ages 21-30), when listening to either sinewave or noiseband vocoded stimuli. The original stimulus was a Spanish sentence, recorded by 40 native speakers of Spanish (20 males, 20 females). All stimuli were manipulated to contain 3- 16 channels, using sinewave and noiseband vocoding methods. The cut-off frequency of the temporal envelope was maintained at 400 Hz for sinewave carriers, and 160 Hz for noiseband carriers. Overall, results showed far better performance with the sinewave vocoded stimuli compared to the noiseband vocoded stimuli, particularly when the number of spectral channels was significantly reduced. The differences in performance can be readily explained, as a modulation of a sinewave generates side bands which convey information about the spectral content of the signal. For this reason, noiseband vocoding may better represent CI processing. Taken together, these studies show that gender and talker identification are difficult tasks among younger listeners when spectral cues are severely limited.

Despite such findings in younger listeners, the influence of spectral degradation on gender identification among older listeners has yet to be investigated. While some studies showed that speech comprehension may be affected by aging when stimuli are spectrally degraded (Peelle & Wingfield, 2005; Schvartz et al., 2008; Sheldon et al., 2008a; 2008b), there is also sufficient reason to speculate that aging affects spectrally-degraded voice pitch processing. For example, Vongpaisal and Pichora-Fuller (2007) showed that periodicity coding of resolved harmonics was slightly worse among older listeners compared to younger listeners when individuals were measured on an F0 discrimination task. Further studies demonstrate senescent decline in the perception of temporal *envelope* periodicity information. For example, Purcell and colleagues (Purcell et al., 2004) measured the envelope following responses in younger and older listeners, and demonstrated reduced amplitude in the older individuals but only for frequencies greater than 100 Hz. The authors hypothesized that such results support decreased temporal acuity in the aging auditory system, and in particular, a deficit in auditory brainstem function. These results are corroborated by subsequent electrophysiologic evidence (Leigh-Paffenroth & Fowler, 2006; Grose et al., 2009) and collectively endorse diminished periodicity coding in elderly listeners, but only at higher

modulation rates (~ 100 Hz). Lastly, neurophysiologic data also supported age-related differences in temporal envelope periodicity coding at the level of the inferior colliculus (Walton et al., 2002), particularly at higher modulation rates. Those modulation rates at which age-related differences in performance are most commonly observed (~ 100 Hz) correspond to those which are important for voice-pitch information and talker gender discrimination.

In summary, it is widely believed that CI users utilize F0 information available in the temporal envelope to determine voice-pitch, and consequently talker-gender. While several studies have measured the ability of younger listeners to identify talker-gender when stimuli are spectrally degraded, it is possible that aging adversely affects this ability. Specifically, some studies provide support for age-related declines in temporal envelope periodicity processing (Leigh-Paffenroth & Fowler, 2006; Grose et al., 2009; Purcell et al., 2004; Walton et al., 2002). It is unknown, however, if these age-related differences obtained using strictly controlled, non-speech stimuli would translate to F0-based gender discrimination performance using a natural speech stimulus. The objective of the present investigation was to measure F0-based gender discrimination in younger and older NH listeners, under conditions of spectral and temporal degradation. It was hypothesized that older listeners would have greater difficulty under conditions with a reduced number of spectral channels due to a greater reliance on temporal envelope cues.

Methods and Materials

Participants

Ten younger (ages 21-28) and ten older (ages 60-73) male and female NH listeners participated in the current investigation. For the purposes of this study, NH was defined as air-conduction audiometric thresholds ≤ 20 dB HL from 250-4000 Hz in the test ear (ANSI, 2004). Average audiometric thresholds and corresponding standard deviations for each group are provided in Table I. These values reflect data obtained in the test ear only. All participants reported to be in good health and denied a significant history of audiologic or otologic disease.

Stimuli

Stimuli were six naturally produced and digitally recorded (sampling rate = 22,050 Hz) vowel tokens (/h/vowel/d/ context, Hillenbrand et al., 1995): /æ/, /i/, /o/, /ɜ/, /e/, /u/. Each of the six recorded vowels was spoken by four male and four female talkers, for a total of 48 vowels (6 vowels $\times$ 8 talkers). Vowels were selected from a larger corpus consisting of 144 vowels, originally recorded at the House Ear Institute (HEI) and were available within the Computer-Assisted-Speech-Training (CAST) software developed by Qian-Jie Fu at HEI (Fu, 2008).

The F0 values were analyzed and manipulated using the autocorrelation method (Boersma, 1993) available in Praat (version 5.0.27) (Boersma & Weenink, 2008). Specifically, the average F0 value of each token was calculated then shifted to match a specified target (within 1 Hz) for a given condition and talker. Assignment of target F0 values was dependent upon the original F0 value of each token; that is, the lowest F0 value among the original tokens was assigned the lowest target F0 value. These F0 targets were selected to create two general sets of stimuli for the current investigation: 'Expanded' and 'Compressed'. For the Expanded condition, F0 values ranged from 100 Hz (lowest F0 of male voice) to 275 Hz (highest F0 value of female voice). For the Compressed condition, the range of F0 values was reduced so that the lowest value was 150 Hz (lowest F0 of male voice) to 220 Hz (highest F0 value of female voice). Each manipulated token was stored on

computer hard drive then uploaded into the CAST program at the time of testing (see below). Table II and Table III provide the target average F0 values (and F0 range) for each talker in the Expanded and Compressed conditions, respectively. Based on these values, average difference in F0s between the male and female talkers was 100 Hz and 40 Hz for the Expanded and Compressed conditions, respectively. It is important to point out that while the average pitch was manipulated, the natural pitch fluctuations within a token remained unaltered.

Noiseband vocoding was accomplished online using TigerCIS within the CAST interface. This was done after each of the pitch manipulations described above. Noiseband vocoding methods used in the current study were comparable to those described in Shannon et al., (1995). Stimuli were first band-pass filtered into 1, 4, 8, 16, or 32 channels depending on the specific condition, using fourth-order Butterworth filters (24 dB/octave).

All stimuli were filtered from 100-4000 Hz, and specific division frequencies for each case (*i.e.* number of channels) were determined based on the logarithmic equation provided by Greenwood (1990). See Table IV for division frequency values. The temporal envelope was then extracted from each frequency band using half-wave rectification and low-pass filtering. The cut-off frequency of the low-pass filter was 20, 50, 100, 200 or 400 Hz, depending on the condition. The original fine structure of each filter was then replaced with corresponding band-pass noise (fourth-order Butterworth filters), and the temporal envelope was used to modulate the noise bands. The modulated noise bands were summed and then normalized to a specified but equal RMS value to result in the final noiseband vocoded speech token.

In order to demonstrate how noise band vocoding manipulates spectral and temporal cues, amplitude spectra and time-amplitude waveforms of the vowel h/ʔ/d are shown in Figures 1 and 2, respectively, with each figure containing varying degrees of spectral resolution (number of channels). In each case the vowel was spoken by the same male talker (F0= 100 Hz), and the cut-off frequency of the low-pass temporal envelope filter is equal to 400 Hz in all cases. As shown in Figure 1, spectral peaks become obscured with fewer spectral channels. Likewise, as shown in Figure 2, the temporal fine structure information is degraded when stimuli are noiseband vocoded, but the temporal envelope periodicity information is available even when spectral cues are absent. In this case, the envelope modulates at the rate of the F0, 100 Hz.

The channel and envelope conditions differed in the Expanded and Compressed stimuli; this was decided based on pilot data that showed the Expanded conditions to be easier than the Compressed conditions. For example, performance reached nearly 100% with 16 channels in the Expanded condition therefore the 32 channel condition was omitted from the test protocol. Channel and envelope conditions were adjusted accordingly within the Expanded and Compressed conditions, thereby resulting in four channel/five envelope conditions for Expanded stimuli, and five channel/four envelope conditions for Compressed stimuli.

Procedure

A two-alternative forced-choice (2AFC) method was used to measure gender identification abilities through the CAST graphical user interface. Using a computer mouse and screen, the participant was instructed to select “start” in order to hear the first vowel token. Following the presentation, two rectangular boxes appeared on the computer screen: one labeled “Male” and the other “Female”. The subject was instructed to identify the gender of the speaker by selecting the appropriate box; there was no time limit for responding. After making a selection the next stimulus was presented 1 second later. This continued until the end of a run, which consisted of 48 vowel tokens (1 repetition × 8 talkers × 6 vowels).

In order to reduce the possibility of order effects, half of the listeners were first presented with stimuli from the Expanded condition, whereas half of the listeners were first presented with stimuli from the Compressed condition. Prior to being tested using the degraded stimuli for each condition, listeners listened to one run of unprocessed stimuli; all listeners were required to achieve 90% accuracy on this task before continuing on in the study. No subject was excluded from the study based on this criterion. Participants were then presented with three presentation cycles, each consisting of a random ordering of all possible noiseband vocoded conditions. Specifically, a cycle consisted of 40 runs (20 Expanded – 4 channel conditions $\times$ 5 envelope conditions plus 20 Compressed – 5 channel conditions $\times$ 4 envelope conditions). Within each of the 40 runs, the 48 vowel tokens were presented in a completely random order. The first presentation cycle served as practice, and an average percent correct was calculated based on the final two cycles.

Stimuli were output through an external soundcard (Edirol 25-UAEX) and mixer (Rane SM26B), before being delivered monaurally through a calibrated circumaural headphone (Sennheiser HDA 200) at a level of 65 dB SPL. For most listeners, stimuli were presented to the right ear with the exception of those listeners ($N=2$) whose right-ear audiometric thresholds did not meet the criteria for normal hearing (as defined under “Participants” above). In this case, stimuli were presented to the better-hearing (left) ear. Testing was performed in two-hour time blocks, with breaks provided at the listener’s discretion. The total duration of the experiment varied from 6-10 hours (3-5 sessions).

Results

Prior to all analyses, percent correct scores were converted to rationalized arcsine units (RAUs) (Studebaker, 1985). If it was found that the assumption of sphericity was violated when conducting an analysis of variance (ANOVA), the Greenhouse-Geisser correction was used to interpret the results.

Overall effects of age on gender identification

Results from the Expanded condition are shown in Figure 3, while results from the Compressed condition are shown in Figure 4. A mixed ANOVA was performed within the Expanded and Compressed conditions, to quantify the interactions between age group, as well as availability of spectral and temporal cues. Subject group (age) served as the between-group factor, whereas number of spectral channels (channels) and cut-off frequency of the low-pass temporal envelope filter (envelope) served as the within-group factors.

As expected, performance was better for Expanded than the Compressed condition. The between-group differences, however, varied within each condition. Overall results suggest that, consistent with previous findings, performance improved as the number of spectral channels and temporal envelope cut-off frequency increased. This general pattern was true for both age groups. Also consistent with previous research (Fu et al., 2004; 2005), results demonstrated a significant interaction between the use of spectral and temporal cues: as the number of spectral channels was decreased, listeners showed greater benefit from increasing temporal envelope information. Perhaps, most importantly the results suggested poorer voice gender discrimination among older compared to younger listeners.

For the Expanded condition, there was a significant main effect of channels [$F(3,54) = 441.84, p < 0.01$], a significant main effect of envelope [$F(2,10, 37.91) = 33.99, p < 0.01$], and a significant interaction between channel and envelope [$F(5.13, 92.34) = 5.33, p < 0.01$]. Further, there was a significant main effect of age [$F(1,18) = 16.89, p < 0.01$], as the average performance of the younger NH group exceeded that of the older NH group. Follow-up testing (using multiple paired t-tests with Bonferroni corrections) was performed for the

significant channel $\times$ envelope interaction. Overall, the results from this analysis showed that there was trade-off between spectral and temporal cues. Specifically, listeners' showed greater benefit from increasing temporal envelope information (i.e., increasing low-pass filter cut-off frequency) as the number of spectral channels was reduced. For example, for the 1-channel condition, performance significantly improved by 16.3 RAUs, on average, from 20 to 400 Hz ($t = -5.94$, $p < 0.001$), whereas there was no effect of temporal envelope cut-off frequency in the 16-channel condition.

The ANOVA results for the Compressed condition show that there was a significant main effect of channels [$F(2.63, 72) = 579.06$, $p < 0.01$], a significant main effect of envelope [$F(3, 54) = 10.63$, $p < 0.05$], and a significant main effect of age [$F(1, 18) = 7.55$, $p < 0.05$]. There was also a significant age $\times$ envelope interaction [$F(3, 54) = 3.40$, $p < 0.05$]. Unlike results for the Expanded condition, the interaction between channels and envelope was not significant. Follow-up tests (one-way ANOVA followed by paired t-test with Bonferroni correction) were used to further examine the interaction between age and envelope, and results revealed that there was a significant main effect of envelope within the YNH group [$F(3, 147) = 13.9$, $p < 0.01$], but not within the older NH group. More specifically, the performance of the younger NH group improved by 3.8 – 5.8 RAUs, on average, as the temporal envelope cut-off frequency was increased from 50 to 200 Hz, from 50 to 400 Hz, and from 100-200 Hz ($p < 0.001$). Although these are not large increases in performance, they are significant, which is certainly meaningful given the similarity of F0 values in the Compressed condition. Lastly, it is important to note that a two-way ANOVA showed no significant differences between expanded and compressed conditions or between age groups when listeners performed the gender identification task.

Talker analyses

In order to further analyze the effects of aging on gender identification, we examined gender identification performance with respect to each of the eight talkers (four male, four female). We expanded the mixed ANOVA analyses (described in the previous section) to include the additional within-group factor of talker-F0. We hypothesized that, the between group differences noted in the initial analyses are dependent upon the individual talkers. It is hypothesized that both age groups were able to better identify talker gender when the F0 of the talker is more extreme (lower F0 for a male, higher F0 for a female). However, we anticipated that the improvement in performance as talker-F0 became more extreme would be less in the older compared to the younger groups.

Results from the Expanded condition are shown in Figure 5, while results from the Compressed condition are shown in Figure 6. Within each of these figures, data plotted in the left and right panels represent results obtained in the younger and older groups, respectively. Within each panel, talker-F0 is shown along the abscissa, while performance on the gender identification task is shown along the ordinate (in percent correct). Analyses within the Expanded condition showed a significant interaction between talker and channel [$F(3.5, 63.5) = 6.35$, $p < 0.01$], and follow-up testing (t-tests with Bonferroni correction) indicated that while performance did tend to improve as the talker-F0 became lower (male) or higher (female) it was also dependent on the number of channels available to the listener. For example, in the 1-channel condition performance systematically improved as the F0 of a male talker decreased (at each 25 Hz decrease from 175 to 100 Hz, $p < 0.05$). Further, the same pattern does not hold true as the number of spectral channels increases and spectral cues become more available: as shown in Figure 5, performance is fairly equivalent regardless of talker-F0 in the 16 channel condition. It is possible, however, that these results could be subject to ceiling effects. Results also indicated a significant interaction between talker and age group [$F(7, 126) = 2.407$, $p < 0.05$]; this overall finding was not dependent on

the number of channels (no significant interaction between talker-F0, age group, and number of channels).

Analyses for the Compressed condition revealed performance patterns similar to those for the Expanded condition. However, in this case the interaction between talker-F0 and age group was not significant. While Figure 6 shows that performance was dependent on talker-F0, this effect did not depend on the age of the listener. This result would suggest that younger listeners are better able to use differences in talker-F0 to improve gender identification performance but only when these values are more diverse.

Taken together, these results are somewhat consistent with our hypothesis that, in some cases, older listeners are not able to use temporal envelope F0 cues to differentiate talker gender to the same extent as their younger counterparts. Specifically, analyses suggest that when F0 values of talkers are largely different from one another (as in the Expanded conditions), younger listeners are better able to use temporal envelope cues to determine if the stimulus was spoken by a male or female compared to older listeners. This effect is clearly shown in Figure 5: while the overall performance of older listeners is poorer than that of the younger listeners, the two age groups exhibit different patterns of performance dependent upon the temporal envelope information. When spectral cues are reduced (Figure 5, 1 channel condition), the performance of younger listeners improves with increasing temporal envelope information (for both male and female talkers). The performance of the older listeners, however, is less dependent on temporal envelope information (particularly for female voices/higher F0 values). This pattern may suggest that, in some cases, older listeners cannot use higher rate temporal envelope information to the same extent as their younger counterparts, even when such cues are made readily available to the listener.

Discussion

In keeping with previous investigations (Fu et al., 2004; 2005), the listeners who participated in the present study demonstrated a systematic relationship between the utilization of temporal and spectral cues when listening in a quiet background. Spectral cues undoubtedly provide better resolution when coding voice-pitch information, and temporal envelope cues are seemingly extraneous when listeners are provided with 16 or 32 spectral channels. However, when spectral cues are absent, some listeners are able to utilize temporal envelope cues to perceive voice-pitch, albeit to a lesser extent than when all spectral cues are available.

The results presented in the current study expand upon those reported in previous investigations measuring temporal envelope processing in aging adults. Previous neurophysiological, electrophysiological, and psychoacoustic studies have alluded to the presence of senescent decline in envelope periodicity processing, while the current study demonstrated that such temporal processing deficits translate to voice pitch coding of natural speech when spectral cues are decreased or absent altogether. Overall, it appears that, while older listeners were able to utilize temporal envelope cues to perceive voice pitch in some cases, the resolution and discrimination of temporal envelope modulation rate was worse than that of their younger NH peers. For example, in the current study we implemented two main conditions: one in which the range of F0 values was large (Expanded), and the second in which the range of F0 values was compressed (Compressed). Based on the results of the Expanded condition, older listeners were able to sometimes discriminate voice gender, even when spectral cues are severely degraded as long as sufficient temporal envelope cues were available. In doing so, they were presumably coding voice pitch using differences in the modulation rates of the residual temporal envelope (see Figure 2D). Results from the Compressed condition, however, revealed that older listeners lack the resolution to

discriminate between residual temporal envelope modulation rates when F0 values (and consequently amplitude modulation rates) are more similar to one another. These findings agree with previous reports of impaired periodicity coding in older listeners for high rate information (Leigh-Paffenroth & Fowler, 2006; Grose et al., 2009; Pichora-Fuller et al., 2007; Purcell et al., 2004; Walton et al., 2002; Walton, 2010).

Although several cues contribute to the perception of talker-gender, further analyses indicated that listeners in the current study were indeed using F0 cues in order to perform the gender identification task. This was particularly true as spectral cues became severely degraded, in which case performance was presumably driven by differences in the modulation rate of the temporal envelope.

Collectively, evidence suggests that CI listeners are able to utilize 8-10 spectral channels, at best (Friesen et al., 2001). Looking at the results for the 8-channel conditions in the current study, it appears that even younger CI users would experience difficulty coding voice-pitch information, as the younger NH participants' performance was 92.5 and 79.4 RAUs for the Expressed and Compressed conditions, respectively when all temporal envelope cues were available (*i.e.*, 400 Hz cut-off). Older listeners, however, exhibited poorer performance in the 8-channel condition when compared to their younger NH peers, as they achieved average scores of 80.9 and 74.1 RAUs for Expanded and Compressed conditions, respectively (*i.e.*, 400 Hz cut-off).

Although little is known about the subtle but potentially significant interactions between aging and cochlear implantation, CIs are undoubtedly beneficial for older listeners (Vermeire et al., 2005; Chatelin et al., 2004). Based on the results of this study (along with a growing number of publications citing temporal envelope processing deficits in older listeners), it is reasonable to conjecture that older CI recipients might be at a disadvantage when distinguishing between talker genders. One caveat to this conclusion is the fact that electrical stimulation of the auditory nerve results in greater sensitivity to temporal amplitude modulations, compared to acoustic stimulation (Shannon, 1992). Another point of debate is the extent to which using CI simulations reflects the performance of older CI users, particularly when relying on temporal cues. Specifically, older adults with hearing loss (CI recipients) may exhibit even greater temporal processing deficits compared to older adults with normal peripheral hearing (Lorenzi et al., 2009). Therefore, it is possible that results of the current study underestimate temporal processing deficits of CI recipients. Lastly, there is little doubt that training and listening experience would improve CI users' performance, but it is impossible to predict how aging would influence performance over time. Even though the results presented in the current study do not reveal an extraordinary effect of age on the spectro-temporal processing of voice-pitch cues, the implication for older CI users could be rather considerable given the importance of F0 coding in daily communication (Arehart et al., 1997; Summers & Leek, 1998; Brungart, 2001).

Conclusion

In summary, the current investigation demonstrated a significant interaction between spectro-temporal envelope processing and the aging auditory system. The results of this study complement those of previous investigations that cited age-related changes in temporal periodicity processing when non-speech stimuli were used. The present findings suggest that such age-related temporal deficits translate to changes in voice-pitch processing of spectrally degraded speech stimuli. Collectively, these findings could have important implications for the rehabilitation of older adult CI users.

Acknowledgments

Many thanks to Qin-Jie Fu, Ph.D. (House Ear Institute) for software support and use of his research interface. We would like to thank the subjects for their time and willingness to participate in the current study. This work was funded by NIH/NIDCD grant no. R01DC004786 to MC, and training grant T32DC000046 to KCS.

Funding acknowledgments: This work was funded by NIH/NIDCD grant no. R01DC004786 to MC, and training grant T32DC000046 to KCS.

References

- ANSI. American National Standard Specification for Audiometers. American National Standards Institute; New York: 2004. ANSI S3.6-2004
- Arehart KH, King CA, McLean-Mudgett KS. Role of fundamental frequency differences in the perceptual separation of competing vowel sounds by listeners with normal hearing and listeners with hearing loss. *J. Speech Hear Lang. Res.* 1997; 40:1434–1444.
- Boersma, P. Proceedings of the Institute of Phonetic Sciences. Vol. 17. University of Amsterdam; 1993. Accurate short-term analysis of the fundamental frequency and the harmonic-to-noise ratio of a sampled sound; p. 97-110.
- Boersma, P.; Weenink, D. [Retrieved: October 1, 2008] Praat: doing phonetics by computer. Version 5.0.342008. [Computer Program]from <http://www.praat.org>
- Brungart DS. Informational and energetic masking effects in the perception of two simultaneous talkers. *J. Acoust. Soc. Am.* 2001; 109:1101–1109. [PubMed: 11303924]
- Chatelin V, Kim EJ, Driscoll C, Larky J, Polite C, Price L, Lalwani A. Cochlear implant outcomes in the elderly. *Otol Neurotol.* 2004; 25:298–301. [PubMed: 15129109]
- Chatterjee M. Modulation masking in cochlear implant listeners: envelope versus tonotopic components. *J. Acoust. Soc. Am.* 2003; 113:2042–2053. [PubMed: 12703715]
- Chatterjee M, Oba SI. Across- and within-channel envelope interactions in cochlear implant listeners. *J. Assoc. Res. Otolaryngol.* 2004; 5:360–375. [PubMed: 15675001]
- Childers DG, Wu K. Gender recognition from speech. Part II: Fine analysis. *J. Acoust. Soc. Am.* 1991; 90:1841–1856. [PubMed: 1755877]
- Dorman MF, Loizou PC, Rainey D. Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *J. Acoust. Soc. Am.* 1997; 102:2403–2411. [PubMed: 9348698]
- Friesen LM, Shannon RV, Baskent D, Wang X. Speech recognition in noise as the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *J. Acoust. Soc. Am.* 2001; 110:1150–1163. [PubMed: 11519582]
- Fu Q-J, Chinchilla S, Galvin JJ. The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *J. Assoc. Res. Otolaryngol.* 2004; 5:253–260. [PubMed: 15492884]
- Fu Q-J, Chinchilla S, Nogaki G, Galvin JJ. Voice gender identification by cochlear implant users: The role of spectral and temporal resolution. *J. Acoust. Soc. Am.* 2005; 118:1711–1718. [PubMed: 16240829]
- Fu, Q-J. [Retrieved: 2008] TigerSpeech Technology: Computer Assisted Speech Training. Version 5.04.022008. [Computer Program]from www.tigerspeech.com
- Goldstein JL. An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustical Society of America.* 1973; 54:1496–1516. [PubMed: 4780803]
- Gonzalez J, Oliver JC. Gender and speaker identification as a function of the number of channels in spectrally reduced speech. *J. Acoust. Soc. Am.* 2005; 118:461–470. [PubMed: 16119365]
- Greenwood DD. A cochlear frequency-position function for several species—29 years later. *J. Acoust. Soc. Am.* 1990; 87:2592–2605. [PubMed: 2373794]
- Grose JH, Mamo SK, Hall JW. Age effects in temporal envelope processing: Speech unmasking and auditory steady state responses. *Ear Hear.* 2009; 30:568–575. [PubMed: 19633565]
- Henry BA, Turner CW. The resolution of complex spectral patterns by cochlear implant and normal hearing listeners. *J. Acoust. Soc. Am.* 2003; 113:2861–2873. [PubMed: 12765402]

- Henry BA, Turner CW, Behrens A. Spectral peak resolution and speech recognition in quiet: normal hearing, hearing impaired, and cochlear implant listeners. *J. Acoust. Soc. Am.* 2005; 118:1111–1121. [PubMed: 16158665]
- Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 1995; 97:3099–3111. [PubMed: 7759650]
- Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* 1990; 87:820–857. [PubMed: 2137837]
- Kovacic D, Balaban E. Voice gender perception by cochlear implantees. *J. Acoust. Soc. Am.* 2009; 126:762–775. [PubMed: 19640042]
- Leigh-Paffenroth ED, Fowler CG. Amplitude-modulated auditory steady-state responses in younger and older listeners. *J. Am. Acad. Aud.* 2006; 17:582–597.
- Licklider JCR. A duplex theory of pitch perception (reproduced in Schubert, 1979, 155–160). *Experientia.* 1951; 7:128–134. [PubMed: 14831572]
- Lorenzi C, Debruille L, Garnier S, Fleuriot P, Moore BC. Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal. *J. Acoust. Soc. Am.* 2009; 125:27–30. [PubMed: 19173391]
- Meddis R, Hewitt MJ. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *Journal of the Acoustical Society of America.* 1991a; 89:2883–2894.
- Meddis R, Hewitt MJ. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity. *J. Acoust. Soc. Am.* 1991b; 91:233–245. [PubMed: 1737874]
- Murry T, Singh S. Multidimensional analysis of male and female voices. *J. Acoust. Soc. Am.* 1980; 68:1294–1300. [PubMed: 7440851]
- Peelle JE, Wingfield A. Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *J Exp Psychol Hum Percept Perform.* 2005; 31:1315–1330. [PubMed: 16366792]
- Pichora-Fuller MK, Schneider BA, MacDonald E, Pass HE, Brown S. Temporal jitter disrupts speech intelligibility: A simulation of auditory aging. *Hear. Res.* 2007; 223:114–1121. [PubMed: 17157462]
- Purcell DW, John SM, Schneider BA, Picton TW. Human temporal auditory acuity as assessed by envelope following responses. *J. Acoust. Soc. Am.* 2004; 116:3581–3593. [PubMed: 15658709]
- Qin MK, Oxenham AJ. Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear Hear.* 2005; 26:451–460. [PubMed: 16230895]
- Ritsma RJ. Frequencies dominant in the perception of the pitch of complex sounds. *Journal of the Acoustical Society of America.* 1967; 42:191–198. [PubMed: 6052077]
- Schvartz KC, Chatterjee M, Gordon-Salant S. Recognition of spectrally degraded phonemes by younger, middle-aged, and older normal-hearing listeners. *J Acoust Soc Am.* 2008; 124:3972–3988. [PubMed: 19206821]
- Shackleton TM, Carlyon RP. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J. Acoust. Soc. Am.* 1994; 95:3529–3540. [PubMed: 8046144]
- Shaddock-Palombi P, Backoff PM, Caspary DM. Responses of young and aged rat inferior colliculus neurons to sinusoidally amplitude modulated stimuli. *Hear Res.* 2001; 153:174–180. [PubMed: 11223307]
- Shannon RV. Temporal modulation transfer functions in patients with cochlear implants. *J. Acoust. Soc. Am.* 1992; 91:2156–2164. [PubMed: 1597606]
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Eklid M. Speech recognition with primarily temporal cues. *Science.* 1995; 270:303–304. [PubMed: 7569981]
- Sheldon S, Pichora-Fuller MK, Schneider BA. Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *J Acoust Soc Am.* 2008a; 123:489–499. [PubMed: 18177176]
- Sheldon S, Pichora-Fuller MK, Schneider BA. Effect of age, presentation method, and learning on identification of noise-vocoded words. *J Acoust Soc Am.* 2008b; 123:476–488. [PubMed: 18177175]

- Schouten JF. The residue and the mechanism of hearing. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Series A*. 1940; 43:991–999.
- Studebaker GA. A 'rationalized' arcsine transform. *J. Speech. Lang. Hear. Res.* 1985; 28:455–463.
- Summers V, Leek MR. F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *J. Speech. Lang. Hear. Res.* 1998; 41:1294–1306. [PubMed: 9859885]
- Terhardt E. Pitch, consonance and harmony. *Journal of the Acoustical Society of America*. 1974; 55:1061–1069. [PubMed: 4833699]
- Vermeire K, Brokx JPL, Wuyts FL, Cochet E, Hofkens A, Van de Heyning PH. Quality of life benefit from cochlear implantation in the elderly. *Otol Neurotol*. 2005; 26:188–195. [PubMed: 15793403]
- von Helmholtz, H. *On the sensations of tone* (English translation A.J. Ellis, 1885, reprinted 1954). Dover; New York: 1877.
- Vongpaisal T, Pichora-Fuller MK. Effect of age on F0 difference limen and concurrent vowel identification. *J Speech Hear Lang Res*. 2007; 50:1139–1156.
- Vongpaisal T, Trehub SE, Schellenberg EG, van Lieshout P, Papsin BC. Children with cochlear implants recognize their mother's voice. *Ear Hear*. 2010; 31:555–566. [PubMed: 20588121]
- Vongphoe M, Zeng FG. Speaker recognition with temporal cues in acoustic and electric hearing. *J. Acoust. Soc. Am*. 2005; 118:1055–1061. [PubMed: 16158660]
- Walton JP, Simon H, Frisina RD. Age-related alterations in the neural coding of envelope periodicities. *J. Neurophysiol*. 2002; 88:565–578. [PubMed: 12163510]
- Wightman FL. The pattern transformation model of pitch. *Journal of the Acoustical Society of America*. 1973; 54:407–416. [PubMed: 4759014]

Figure 1. Amplitude spectra for the vowel h/ɪ/d spoken by a male with an average F0=100 Hz. As indicated in the legend, each line represents varying degrees of spectral resolution (number of spectral channels).

Figure 2.

Time-amplitude waveforms for the vowel h/ɛ/d spoken by a male with an average $F_0=100$ Hz. Each panel (2A-2D) represents the same token processed with varying degrees of spectral cues (A= Unprocessed, B= 16 channels, C= 8 channels, D = 1 channel). The cut-off frequency of the low-pass temporal envelope filter is equal to 400 Hz in panels B, C and D. The small insert within each panel represents a close examination of the temporal fine structure within each waveform.

Figure 3.

Results are shown for performance on the gender identification task for the Expanded condition. Each graph represents results obtained within each group (panel 'A' = younger listeners and panel 'B' = older listeners) for each condition, with the number of channels represented along the abscissa and percent correct along the ordinate. The label "ALL" refers to performance using unprocessed stimuli. The various symbols represent the value of the cut-off frequency of the low-pass temporal envelope filter (see legend). The error bars represent ± 1 standard deviation from the mean. Chance performance (50%) is indicated by the dashed horizontal line within each graph.

Figure 4.
Identical to Figure 3, but represents data obtained in the Compressed conditions.

Figure 5.

Gender identification performance is shown according to each individual talker for the Expanded condition. The panels on the left and right show data from the younger and older listeners, respectively. The top panels show data obtained when stimuli were processed to contain 1-channel, and the lower panels show data obtained as more spectral channels were made available to the listener. Within each graph, the thin lines (open symbols) represent varying degrees of temporal envelope cues (see legend), while the thick line (closed symbol) represents average performance collapsed across all temporal envelope conditions. The horizontal dashed line represents chance performance (50%). Error bars represent $\pm$ SE from the mean.

Figure 6.
Identical to Figure 5, but represents data obtained in the Compressed conditions.

Average audiometric pure-tone thresholds for each frequency are provided for each age group (Mean), with the corresponding standard deviation provided below (SD). The pure-tone average (PTA) is also provided. The reported values represent data from participants' test ear. The asterisk refers to frequencies (6000 and 8000 Hz) that were tested, however were not used to determine inclusion into the study, as the experimental stimuli contained information up to 4000 Hz only.

Table 1

Audiometric data for each group											
PTA		250	500	1000	2000	3000	4000	6000*	8000*		
YNH	Mean	6.33	9.5	7	6.5	5.5	7.5	3.5	7.5	6.5	
	SD	3.58	5.98	4.83	4.11	5.98	5.97	4.74	4.26	5.29	
ONH	Mean	10.5	15	8.5	12	11	13.5	16.5	19.5	27	
	SD	1.93	5.27	4.11	4.21	3.16	3.37	4.11	8.31	16.36	

Table II

Target average fundamental frequency (F0) values and actual ranges (Hz) estimated for each talker, for the Expanded condition. Actual average F0 values for each vowel stimulus may have varied by ± 1 Hz for each talker.

F0 values (Expanded)														
F0	h/ɪ/d		h/ʊ/d		h/ɜ/d		h/e/ə/d		h/ɔ/ɒ/d		h/ɑ/d		h/ɹ/d	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
100	98	102	97	105	96	104	95	106	98	103	98	104		
125	120	138	123	128	122	128	121	128	122	128	124	127		
150	144	155	146	152	147	154	145	154	144	156	144	153		
175	169	185	170	182	170	185	170	189	171	178	172	188		
200	190	222	182	225	182	227	183	232	180	238	185	234		
225	209	243	219	230	220	235	213	245	212	248	210	241		
250	240	268	242	262	245	265	242	263	243	261	241	265		
275	237	290	238	286	264	300	260	285	269	283	270	290		

Table III

Identical to Table II, but representative of stimuli used in the Compressed condition.

F0 values (Compressed)													
F0	h/ɹ/d		h/ɔ/d		h/ʌ/d		h/e/ɘ/d		h/ɛ/ɜ/d		h/ɪ/ɪ/d		h/ɹ/d
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min Max
150	148	151	147	154	144	154	144	157	147	152	148	153	
160	154	164	157	162	157	164	157	162	156	163	159	161	
170	165	176	167	172	165	174	164	173	164	178	163	174	
180	175	190	175	187	175	193	178	192	176	185	177	194	
190	180	224	171	217	170	220	171	224	169	230	175	226	
200	182	221	192	204	195	210	187	226	186	224	183	220	
210	200	229	201	222	204	225	201	224	203	221	201	226	
220	191	236	182	241	175	255	160	256	172	250	163	247	

Table IV

Division frequencies used for the noise-band vocoding processing. The division number is listed in the first column, with the corresponding condition (number of channels) listed towards the top of the table.

Division frequencies for noise-band vocoding filters (Hz)				
Division #	Number of channels			
	32	16	8	4
1	123	147	204	352
2	147	204	352	863
3	174	272	563	1900
4	204	352	863	
5	236	448	1291	
6	272	563	1900	
7	310	700	2766	
8	352	863		
9	398	1058		
10	448	1291		
11	503	1568		
12	563	1900		
13	629	2295		
14	700	2766		
15	778	3329		
16	863			
17	957			
18	1058			
19	1169			
20	1291			
21	1424			
22	1568			
23	1727			
24	1900			
25	2088			
26	2295			
27	2520			
28	2766			
29	3035			
30	3329			
31	3650			