

Fundamental-frequency discrimination using noise-band-vocoded harmonic complexes in older listeners with normal hearing

Kara C. Schwartz-Leyzac^{a)} and Monita Chatterjee^{b)}

Department of Hearing and Speech Sciences, University of Maryland, 0100 LeFrak Hall, College Park, Maryland 20742, USA

(Received 17 November 2014; revised 6 August 2015; accepted 20 August 2015; published online 30 September 2015)

Voice-pitch cues provide detailed information about a talker that help a listener to understand speech in complex environments. Temporal-envelope based voice-pitch coding is important for listeners with hearing impairment, especially listeners with cochlear implants, as spectral resolution is not sufficient to provide a spectrally based voice-pitch cue. The effect of aging on the ability to glean voice-pitch information using temporal envelope cues is not completely understood. The current study measured fundamental frequency (f_0) discrimination limens in normal-hearing younger and older adults while listening to noise-band vocoded harmonic complexes with varying numbers of spectral channels. Age-related disparities in performance were apparent across all conditions, independent of spectral degradation and/or fundamental frequency. The findings have important implications for older listeners with normal hearing and hearing loss, who may be inherently limited in their ability to perceive f_0 cues due to senescent decline in auditory function.

© 2015 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4929938>]

[EB]

Pages: 1687–1695

I. INTRODUCTION

Several cues aid in one's ability to perceive a target message in the presence of extraneous distracters, but adequate processing of fundamental frequency (f_0) information is critical to communicating in such environments. Voiced speech can be approximated by a harmonic complex signal, with the perceived pitch of listeners' voices roughly corresponding to the f_0 of the harmonic complex. Individuals not only use voice pitch information in speech perception tasks and linguistic processing, including speech intonation recognition (Lehiste, 1970), lexical tone recognition (Chao, 1968), and talker-gender identification (Titze, 1989), but also to separate competing sound sources (Brox and Nootebohm, 1982; Brungart, 2001) and to determine speaker authenticity and voice emotion (Drolet *et al.*, 2014).

A. Effects of aging on perception of periodicity cues

Young normal hearing listeners are quite good at discriminating f_0 s, and typically require approximately 1% difference, or less, between the f_0 of two signals in order to achieve good discrimination. Older normal-hearing listeners, however, demonstrate a reduced ability to discriminate between f_0 s and typically require twice the difference between f_0 s to achieve a performance level equivalent to their younger, normal hearing peers. These findings have

been demonstrated using harmonic complex stimuli and synthetic vowels (Moore and Peters, 1992; Vongpaisal and Pichora-Fuller, 2007).

Similar findings have been reported when electrophysiologic and neurophysiologic measures were used, generally indicating senescent declines in temporal coding. For example, Purcell *et al.* (2004) measured the envelope-following responses in younger and older listeners, and demonstrated reduced amplitude in the older individuals for envelope frequencies greater than 100 Hz. The authors suggested 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 (Anderson *et al.*, 2011; Grose *et al.*, 2009) and collectively demonstrate diminished periodicity coding in elderly listeners, particularly at higher modulation rates (those ≥ 100 Hz).

B. Hearing loss, cochlear implants (CIs) and f_0 perception

Less is known about how hearing loss, and particularly cochlear implantation, may affect an older listener's ability to perceive f_0 cues in the voice pitch range. While individuals with normal hearing rely heavily on f_0 cues for various aspects of speech perception, previous studies have documented that, owing to poorer spectral resolution, listeners with hearing impairment or listeners with CIs are not able to take advantage of f_0 cues as readily as those with normal hearing. These listeners are, however, still able to achieve some, albeit reduced, perception of f_0 cues (Arehart, 1994; Chatterjee and Peng, 2008; Qin and Oxenham, 2005). While sensorineural hearing loss typically results in some

^{a)}Current address: Kresge Hearing Research Institute, Hearing Rehabilitation Center, Department of Otolaryngology, University of Michigan, 1150 West Medical Center Drive, Ann Arbor, MI, 48103, USA. Electronic mail: kleyzac@med.umich.edu

^{b)}Current address: Boystown National Research Hospital, 555 N. 30th Street, Omaha, NE 68104, USA.

broadening of the peripheral auditory filters, most listeners with hearing impairment are able to use low-frequency spectro-temporal cues to process voice pitch. For instance, even listeners with significant bilateral hearing loss who wear a CI on one side, are able to benefit from acoustic f_0 cues transmitted by the hearing aid on the other side (Brown and Bacon, 2010; Kong *et al.*, 2005). In listeners with a moderate to severe hearing loss (who presumably have broadened auditory filter bandwidths) temporal envelope cues may be available via the unresolved harmonics even though spectral cues are diminished (Bernstein and Oxenham, 2006). As spectral cues become degraded, such listeners may rely more on available temporal envelope cues to perceive f_0 information.

Unlike listeners with hearing impairment who are able to use both resolved and unresolved harmonics (via the temporal envelope) to some extent, listeners with CIs rely more on temporal envelope cues, which are somewhat preserved in most speech processing strategies. CIs provide relatively coarse spectral resolution due to electrical current spread and, possibly, additional spectral degradation due to neural degeneration. Furthermore, the temporal fine structure cues required to determine f_0 are not conveyed in CI processing. Previous studies estimate that the average CI users' performance does not improve beyond eight spectral channels, depending on the speech stimulus and listening environment (Fishman *et al.*, 1997; Friesen *et al.*, 2001).

Despite limited spectral cues, it appears that listeners with CIs are able to perceive f_0 cues via the temporal envelope. For example, Chatterjee and Peng (2008) measured modulation frequency discrimination abilities in listeners with CIs using direct stimulation methods. Results were highly variable across subjects but showed median modulation frequency discrimination thresholds (percent Weber fractions) of CI users were approximately 10% to 20% for standard rates of 100 and 200 Hz, respectively. These results approximate thresholds obtained in normal hearing individuals tested with sinusoidally amplitude modulated (SAM) noise stimuli (Formby, 1985; Grant *et al.*, 1998) or noise-band vocoded (NBV) CI simulations (Qin and Oxenham, 2005), despite significant variability in performance among the CI listener cohort. Chatterjee and Peng (2008) showed that performance on the modulation frequency discrimination task correlated with CI listeners' sensitivity to an acoustic f_0 in a speech intonation task when listening with their everyday speech processor. Similarly, Luo *et al.* (2008) showed significant correlations between Mandarin-speaking listeners' lexical tone recognition and their sensitivity to temporal envelope cues. These results suggest that the measured psychophysical sensitivity to temporal envelope cues might translate to a real-life benefit for CI patients. Chatterjee and Peng (2008) and Chatterjee and Oberzutz (2011) found that factors such as the stimulating electrode, the envelope modulation depth, the level, and the availability of loudness cues might influence CI listeners' sensitivity to envelope periodicity cues to some extent. Thus, one might expect that speech processing strategies and processor settings might influence listeners' performance in real-world tasks with multi-channel stimuli. In a recent study, Galvin

et al. (2015) showed that multi-channel and single-channel psychophysical sensitivity to envelope periodicity was similar as long as the overall loudness remained constant. This might partially explain why results obtained in single-channel stimulation paradigms have been found to relate to real-world performance by CI listeners. Despite these recent studies, little is known about how the age of the CI listener might interact with their sensitivity to envelope periodicity. By focusing on this question in younger and older normally hearing listeners attending to CI-simulated stimuli, we address the question independently of the many sources of variability that might confound a study with actual CI patients and/or listeners with hearing loss.

C. Aging and perception of temporal envelope-based pitch

The perception of temporal envelope pitch cues has been quantified among younger normal hearing adult listeners, but less is known about how such abilities might be affected by the age of the adult listener. The ability of older adult listeners to detect the presence of amplitude modulation in broadband noise is relatively unaffected by age (Takahashi and Bacon, 1992). The influence of aging on the ability of adults to detect changes in the frequency of audible amplitude modulation, however, remains unclear. Moore and Peters (1992) investigated the ability of a small group of younger and older normal hearing listeners and older listeners with hearing impairment to detect changes in f_0 , while manipulating the number of harmonics in a 12-harmonic complex. In one condition, listeners were tested on their ability to detect changes in f_0 when stimuli only contained the highest harmonics (6–12); in this case, it was assumed that any discrimination was based on temporal envelope pitch via unresolved harmonics. The performance of older listeners was worse than that of younger listeners, regardless of hearing status, for all reference f_0 values (50–400 Hz). Mean f_0 DLs were approximately 1% in younger normal hearing listeners and ranged from 1.5%–3% for older normal hearing listeners (for 50–400 Hz, respectively). Despite these results, it is unknown if listeners were relying solely on temporal envelope pitch information to perform the task, as a low-frequency masker was not used. Studies have shown that a low-frequency masker is needed to limit the perception of distortion products that aid in the salience of f_0 perception (Smalt *et al.*, 2012). Furthermore, studies suggest that perhaps up to the tenth harmonic of a complex signal may be considered “resolved harmonics” (Bernstein and Oxenham, 2003).

More recently, Souza *et al.* (2011) measured f_0 DLs in younger normal hearing listeners and older listeners with slight to mild hearing loss (gain frequency shaped stimuli were used to account for between-group threshold differences). Stimuli consisted of an eight-channel, noise-band-vocoded (NBV) vowel stimulus. Stimulus f_0 s ranged from 100 to 130 Hz. Results showed that older listeners required larger f_0 DLs compared to younger listeners for both unprocessed and eight channel NBV stimuli. The authors note an apparent interaction between age and stimulus type (the

between-group difference in performance was especially apparent for the NBV stimulus and older adults showed greater variability with the NBV stimulus as well), but this difference was not statistically significant.

In a previous study we measured the ability of younger and older normal hearing listeners to discriminate between voice genders when vowel stimuli were systematically manipulated to contain varying degrees of spectral and temporal cues (Schvartz and Chatterjee, 2012). Results showed that older adults demonstrated poorer results on several conditions, particularly those in which only temporal envelope cues were available (i.e., one-channel condition) to glean voice gender. However, it is uncertain to what degree voice gender cues other than f_0 that are present in speech contributed to these results. Studies have demonstrated that vocal tract length also contributes to the perception of gender identification when spectral cues are reduced (see Fuller *et al.*, 2014, for review).

Stimuli used in Souza *et al.* (2011) were confined to eight-channel NBV processing using an f_0 in the range of 100–130 Hz. It is of interest to systematically measure f_0 discrimination using varying degrees of spectral cues and to do so using f_0 values that reflect both male and female voices, particularly given that previous studies have suggested that age-related differences in performance are greater at higher f_0 values (Grose *et al.*, 2009; Purcell *et al.*, 2004).

Given the importance of complex pitch cues in everyday speech and audio perception, it is important to determine the extent to which aging might affect one’s ability to discriminate between different f_0 s when spectral cues are altered. While previous studies do suggest that aging affects the use of primarily temporal envelope cues for f_0 discrimination, the current study used controlled stimuli to systematically vary available spectral cues. The aim of the current study was to gain a more detailed understanding of how adult age affects f_0 coding as listeners transition from using primarily spectral (or spectro-temporal) to primarily temporal-envelope cues.

II. METHODS

A. Participants

Participants were 25 normal hearing males and females, recruited for placement in two different groups based on their age at the time of testing: younger (ages 21–26; mean = 21.9, standard deviation [SD] = 1.62) or older (ages 60–77; mean = 64.7, SD = 5.43). The younger normal hearing group consisted of 12 individuals (seven males, five females) and the older normal hearing group consisted of 13 individuals (three males, ten females). All participants were required to have pure-tone thresholds ≤ 20 dB hearing level (HL) from 250 to 4000 Hz in the test ear (ANSI, 2004). See Table I for audiometric data. All participants in the older age group obtained scores within the normal range on the Mini Mental State Examination (MMSE) (Folstein *et al.*, 1975), a screening test used to identify gross cognitive dysfunction.

TABLE I. Average audiometric air-conduction pure-tone thresholds (dB HL) at each frequency tested and pure-tone average (PTA = average thresholds at 500, 1000, and 2000 Hz) for each age group. Standard deviations are shown below each respective average value.

		Audiometric pure-tone thresholds (dB HL)							PTA
		Frequency (Hz)							
		250	500	1000	1500	2000	3000	4000	
YNH									
AVG		8.3	8.3	7.1	6.0	7.1	9.5	5	7.5
SD		6.5	4.4	5.8	3.2	4.9	5.2	3.6	4.1
ONH									
AVG		13.8	11.5	11.5	10.8	11.9	15	13.0	11.6
SD		5.8	3.7	3.1	5.1	4.8	4.7	5.2	2.5

B. Stimuli

All signals were created online (44 100 Hz sampling rate) and delivered through a custom graphical user interface developed in MATLAB. Stimuli were harmonic complexes, 300 ms in duration. The starting phase for the f_0 and associated partials was always fixed at 0° . Signals were created by first generating equal-amplitude harmonics between 100 and 4000 Hz. The f_0 and number of harmonics in each signal varied depending upon the condition; however the highest harmonic was selected so that each stimulus did not contain information above 4000 Hz. For each signal, the f_0 and partials were summed to create a harmonic complex series. Noise-band vocoding was similar to the process described by Shannon *et al.* (1995). Each harmonic series was first band-pass filtered into 1, 8, or 24 channels (Chebychev, 40 dB/octave) depending on the condition, and the unaltered temporal envelope was then extracted from each channel using a Hilbert transform. The division frequencies used for the band-pass filtering were determined by using the logarithmic equation provided in Greenwood (1990) which estimates filter characteristics in the cochlea. Specific values of the division frequencies and bandwidths for each filter are shown in Table II. Freshly generated noise was used in each trial to create a noise-band equal in bandwidth and center frequency. Each of these noise-bands was multiplied with the corresponding temporal envelope. The outputs of all of the modulated noise bands were summed and low-pass filtered at 4000 Hz (Chebychev, 80 dB/octave). A Tukey window (a rectangular window with symmetric cosine shaped rising and falling edges) using the MATLAB “tukeywin” command was applied to the time domain of all stimuli to avoid transients and minimize spectral splatter. The rising and falling portions of the window each occupied 10% of the full length of the signal. Across-channel processing delays due to filtering were compensated for using zero-phase forward/backward filters. Last, all stimuli were equalized in root-mean-square value before presentation. Figures 1 and 2 show time amplitude waveforms of the noise-band vocoded harmonics complexes in the case of $f_0 = 100$ and 200 Hz, respectively.

C. Procedure

A two-down, one-up, 3-alternative forced choice (AFC) procedure was used to measure f_0 discrimination

TABLE II. Division frequencies used for noise-band vocoding of harmonic complex stimuli. The minimum (Min), maximum (Max) and bandwidth (BW) frequency values are given for each channel number in the 24- and 8-channel conditions. The 1-channel condition is not shown.

Channel number	Division frequencies for noise-band vocoding (Hz)					
	Number of channels					
	24			8		
Min	Max	BW	Min	Max	BW	
1	100	131	31	100	204	104
2	131	165	34	204	352	148
3	165	204	39	352	563	211
4	204	248	44	563	863	300
5	248	297	49	863	1291	428
6	297	352	55	1291	1900	609
7	352	414	62	1900	2766	866
8	414	484	70	2766	4000	1234
9	484	563	79			
10	563	652	89			
11	652	751	99			
12	751	863	112			
13	863	989	126			
14	989	1131	142			
15	1131	1291	160			
16	1291	1470	179			
17	1470	1672	202			
18	1672	1900	228			
19	1900	2155	255			
20	2155	2443	288			
21	2443	2766	323			
22	2766	3130	364			
23	3130	3539	409			
24	3539	4000	461			

(threshold = 70.7%; Levitt, 1971), in which two of the intervals contained a reference stimulus with f_0 value equal to 100 or 200 Hz, while a third interval contained the experimental value (always greater than the reference f_0). The experimental stimulus was presented at random in one of the three intervals. The f_0 value of the experimental stimulus was adapted for a maximum of ten reversals or 55 total trials (whichever occurred first). A minimum of eight reversals was required to calculate the average f_0 value; if eight reversals could not be reached in 55 trials, the run was aborted. Initial and final adaptive step sizes varied depending on the condition and listener's sensitivity, and were either 4 and 2 Hz, 2 and 1 Hz, or 1 and 0.5 Hz. In general, smaller step sizes were used for a lower reference value (i.e., 100 Hz) and conditions in which stimuli contained more spectral cues (e.g., unprocessed or 24 channels), whereas larger step sizes were required for a higher reference value (i.e., 200 Hz) and conditions in which stimuli contained fewer spectral cues (e.g., 1 channel). The mean was calculated from the last four reversals of a run, and that value was taken as the f_0 DL.

Stimuli were output through an external soundcard [Edirol 25-UAEX (Roland Corporation US, Los Angeles, CA)] and mixer [RaneSM26B (Rane Corporation, Mukilteo, WA)], before being delivered through a calibrated circumaural headphone [Sennheiser HDA 200 (Sennheiser Electronic Corporation, Old Lyme, CT)]. All stimuli were delivered

monaurally at 65 dBA in the better-hearing ear. When both ears were deemed equally sensitive, the right ear was used. Participants were tested in a double-walled, sound-proof booth using a computer interface. The computer interface displayed boxes (labeled "1," "2," and "3") which appeared sequentially, but simultaneous with the corresponding reference or experimental stimuli. Using a mouse, listeners were asked to click on the box that contained the "different" sound (experimental f_0). The inter-stimulus-interval was 400 ms, and there was no time limit within which the subject was asked to respond. After the subject made a selection, the next sequence of sounds was played 600 ms later.

Subjects received practice before being tested; one run of every condition was provided as practice and individuals received feedback about their response in the form of text that appeared at the top of the screen ("correct" or "incorrect"). A practice run of each condition (each reference f_0 for each NBV condition) was presented prior to testing. The practice conditions were presented in a randomized order. After completing the practice run, listeners completed one threshold estimation run in each condition, with conditions run in random order, and this sequence of conditions was completed twice. Feedback was not provided during the test trials. These two runs were used to calculate the final mean performance. In cases when the difference between the thresholds of the two means was greater than 15%, another run was performed, and then the average of all three runs was calculated as the final mean.

III. RESULTS

Results of the f_0 DL task are shown in Figs. 3 and 4 for reference f_0 s of 100 and 200 Hz, respectively. Within each graph, the abscissa represents the degree of spectral degradation/number of channels, whereas the ordinate represents the discrimination threshold in Weber fraction (Δf_0 /reference $f_0 \times 100$) on a log scale. For each box plot shown, the bottom of the shaded area represents the 25th percentile, a line within the box marks the median, and the top of the shaded area indicates the 75th percentile. Whiskers (error bars) above and below the box indicate the 90th and 10th percentiles. Individual data points show outliers, above and below the 90th and 10th percentiles, respectively. The results of younger participants (darker gray shaded boxes) are shown in comparison to the older participants (lighter shaded boxes).

Data were analyzed using SPSS (version 22.0). A split-plot analysis of variance (ANOVA) was performed to analyze the data obtained for the f_0 DL task with two within-group factors (number of channels ["channels"] and reference f_0 value [" f_0 "]), and one between-group factor (Age group ["age"]). All data analyses were conducted on log-transformed values of Weber fractions. The Greenhouse-Geisser correction was used for interpretation of the results when the assumption of sphericity was violated. Analyses revealed a significant main effect of channels ($F[3, 69] = 304.79, p < 0.001$) and a significant main effect of age ($F[1, 23] = 14.75, p < 0.001$). Analyses also indicated a significant interaction between channels and f_0 ($F[3, 69] = 24.97, p < 0.001$). The main effect of Frequency ($F[1,$

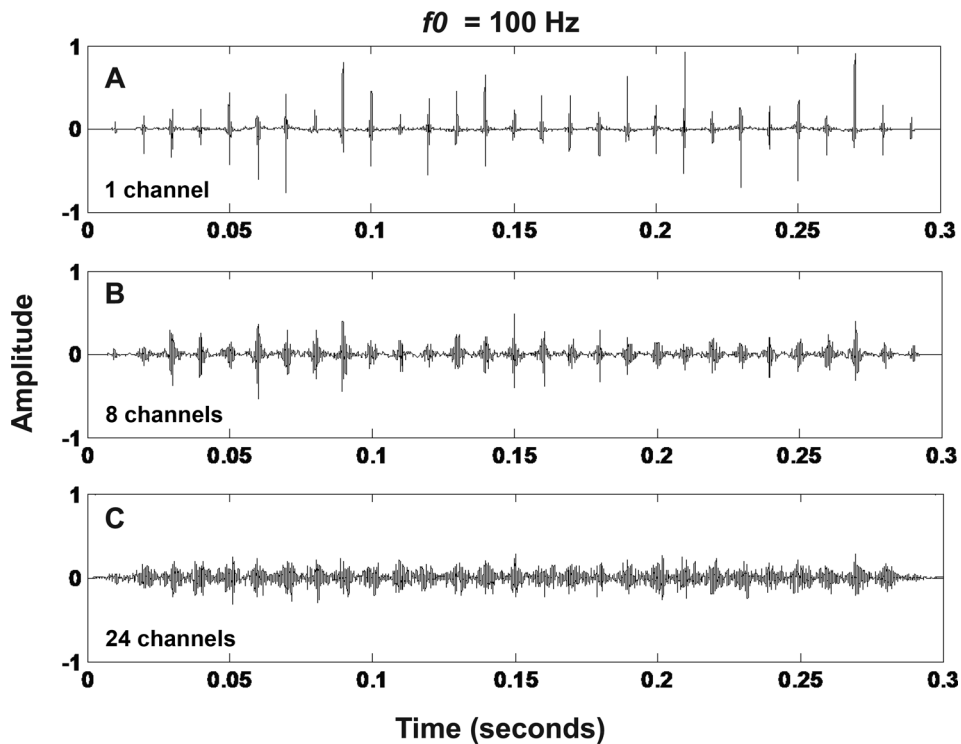


FIG. 1. Time amplitude waveforms of noise-band vocoded harmonic complex stimuli for an $f_0 = 100$ Hz. Shown are the outputs when (A) 1, (B) 8, or (C) 24 division bands are used to create noise-band vocoded stimuli.

23] = 0.00, $p = 0.994$) and interactions between frequency and age ($F[1,23] = 0.482$, $p = 0.494$), number of channels and age ($F[3,21] = 1.45$, $p = 0.255$), and frequency \times number of channels \times age ($F[3, 21] = 0.414$, $p = 0.745$) were not significant. Follow-up analyses included paired or independent t tests and the Bonferroni correction was applied appropriately to interpret the results.

Follow up testing revealed that (collapsed across subject group) within the 100 Hz condition, only the 1- and 24-channel conditions were not significantly different from one

another (Family-wise criteria = $p < 0.05$; Bonferroni corrected value = $p < 0.008$). In all other comparisons, there was a significant effect of number of channels. Both age groups performed best when listening to an unprocessed signal (across-group geometric mean = 1.53 Hz), and performance was worst when listening to the eight-channel condition (across-group geometric mean = 13 Hz). Similarly, within the 200 Hz condition, only the one- and eight-channel conditions were not significantly different from one another (Familywise criteria = $p < 0.05$; Bonferroni corrected

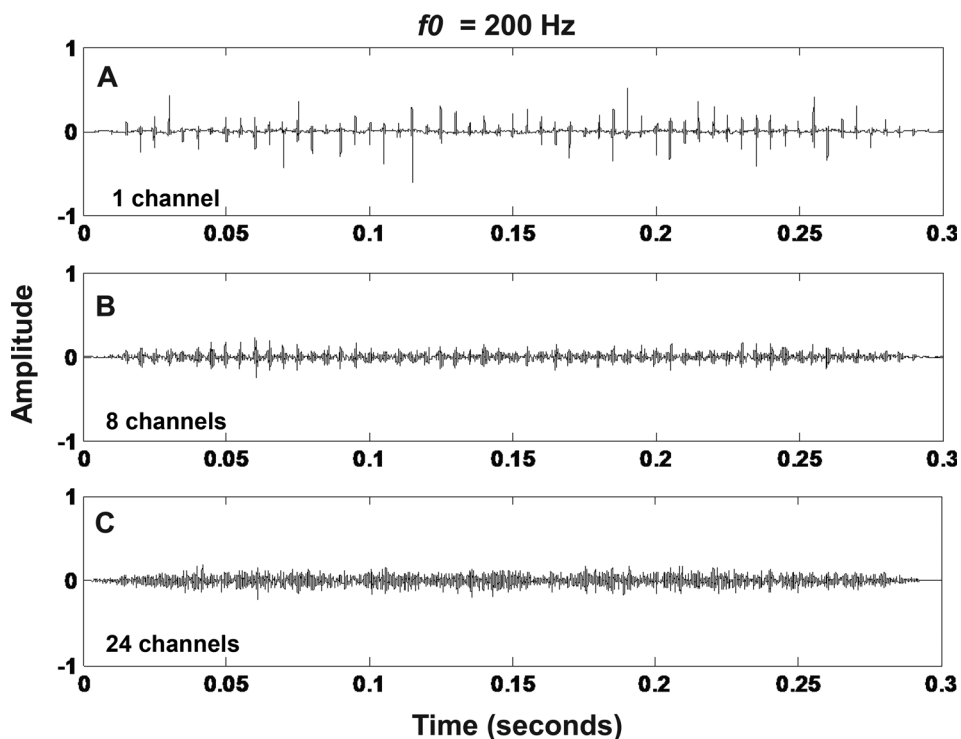


FIG. 2. Same as Fig. 1, but with f_0 of the harmonic complex = 200 Hz.

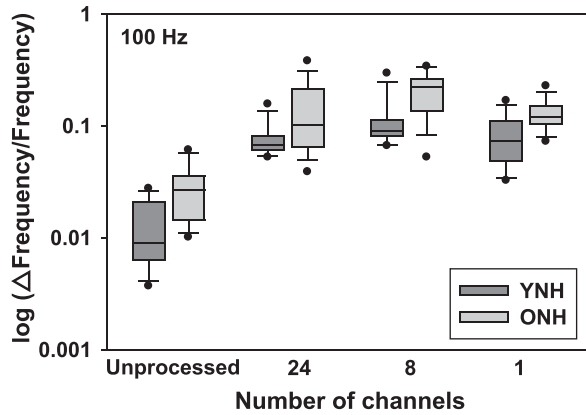


FIG. 3. Results of the fundamental frequency (f_0) discrimination task (Reference = 100 Hz) representing data obtained in younger and older listeners. The abscissa represents the degree of spectral degradation/number of channels, whereas the ordinate represents the discrimination threshold in Weber fraction ($\Delta f_0/\text{reference } f_0 \times 100$) on a log scale. For each box plot shown, the bottom of the shaded area represents the 25th percentile, a line within the box marks the median, and the top of the shaded area indicates the 75th percentile. Whiskers (error bars) above and below the box indicate the 90th and 10th percentiles. Individual data points show outliers, above and below the 90th and 10th percentiles, respectively. The results of younger participants (darker gray shaded boxes) are shown in comparison to the older participants (lighter shaded boxes).

value = $p < 0.008$). In all other comparisons, there was a significant effect of number of channels. Analysis of the 200 Hz reference f_0 condition showed that both groups performed best when listening to an unprocessed signal (across-group geometric mean = 1.15 Hz), and performance was worst when listening to the one- and eight-channel conditions (across-group average geometric means = 1.50 and 1.94 Hz, respectively). When data are collapsed across age group and reference f_0 , follow-up testing revealed that only the one- and eight-channel conditions were not significantly different from one another (Family-wise criteria = $p < 0.05$; Bonferroni corrected value = $p < 0.008$). Follow-up testing also revealed that, in general, performance for the 100 Hz reference frequency yielded better difference limens when compared to performance for the 200 Hz condition (collapsed across age group and number of channels). The only exception to this finding was a non-significant difference between the one- and eight-channel conditions.

Subjects were recruited to have normal hearing through 4000 Hz, and stimuli were carefully constructed to keep within the limits of normal hearing for all subjects. However, correlational analyses were performed in order to address possible contributions of peripheral hearing sensitivity to the results of the f_0 discrimination task. Pearson correlation (two-tailed) analyses were performed between the average pure-tone audiometric thresholds from 250 to 4000 Hz and f_0 DL values for each test condition (eight comparisons total), for all subjects. A Bonferroni correction was applied as is appropriate when examining multiple comparisons. Results showed no significant relationship between the puretone thresholds and f_0 DL limens for any of the eight comparisons (Family-wise criteria = $p < 0.05$; Bonferroni corrected value = $p < 0.006$). However, results should be interpreted with caution due to the limited number of subjects who participated in the study.

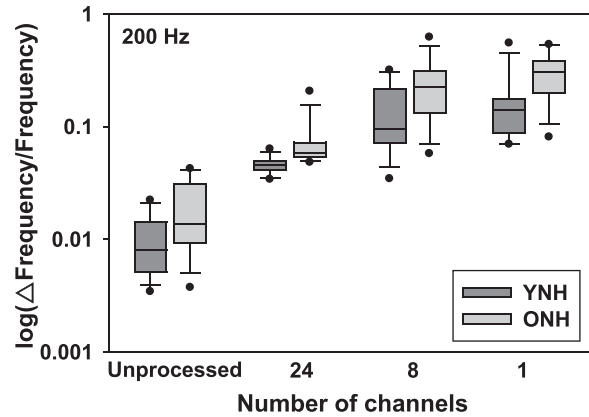


FIG. 4. Same as Fig. 3, with a f_0 reference rate = 200 Hz.

IV. DISCUSSION

Taken together, the present results show that the performance of the older normal hearing group was worse than that of the younger normal hearing group for all conditions. These findings suggest impaired periodicity coding in older listeners, regardless of the degree to which spectral cues were degraded. In other words, older listeners' ability to perceive f_0 information is not dependent on the mechanism with which cues are conveyed (via the temporal envelope, temporal fine structure cues or a combination of these two mechanisms).

Previous studies using non-speech (harmonic complexes) and synthetic speech stimuli (synthetic vowels) revealed that f_0 DLs among older listeners were $\sim 2\%$ whereas those of younger listeners were $\sim 1\%$ when harmonics could be resolved (Moore and Peters, 1992; Vongpaisal and Pichora-Fuller, 2007). The current study corroborates such findings, as average f_0 DLs for unprocessed harmonic complexes were 1.17% (100 Hz) and 0.93% (200 Hz) for younger normal hearing listeners. Conversely, average f_0 DLs for unprocessed harmonic complexes were 2.79% (100 Hz) and 1.9% (200 Hz) for older normal hearing listeners. The effect of aging on periodicity coding was independent of modulation frequency for all stimuli, as similar age-related deficits were found for 100 and 200 Hz reference stimuli.

For the current study, a consideration of the cues available to the listener in each condition may shed some light on the patterns of observed results. The one-channel condition yielded deeper modulation than the eight-channel condition, because of the greater number of harmonics being summed. Further, it is likely that the presence of auditory filters in the eight-channel condition reduces the effective modulation depth when compared to the modulation depth of the one-channel condition. The 8-channel condition would have likely provided poorer envelope cues within each band than the 1-channel condition, and also poorer spectral cues for pitch than the 24-channel condition. The 24-channel condition would have provided the shallowest modulation, but better spectral resolution. This pattern of cue-availability is observed in the perceptual data in the 100-Hz f_0 condition, where listeners' thresholds for 1-channel and 24-channel

stimuli were similar, and lower (better) than their thresholds for eight-channel stimuli. In the 200-Hz f_0 condition, listeners' envelope sensitivity may have been too poor for them to benefit from the deeper modulation depths in the one-channel condition over the eight-channel condition: thus, the thresholds obtained in those two conditions were equivalent and both higher than in the 24-channel condition. One caveat to this interpretation and its application to CI recipients is that the 24-channel noise-band vocoding simulation provides different (and most likely better) spectral cues than those available to most CI recipients.

Modern CI devices convey approximately eight to ten spectral channels on average (Friesen *et al.*, 2001; Fishman *et al.*, 1997). The results of the present study show that when listening with eight spectral channels, younger listeners' average difference limen was approximately 11% and 12.5% for 100 and 200 Hz reference f_0 conditions, respectively. Likewise, when listening with eight spectral channels, older listeners' average difference limen was approximately 20% and 23% for 100 and 200 Hz reference f_0 conditions, respectively. This comparison suggests that, on average, older CI recipients may require a greater difference in f_0 in order to reliably discriminate between two sequential talkers.

These results obtained in younger normal hearing listeners closely match those reported in previous studies (Qin and Oxenham, 2005). In the current study, older listeners exhibited greater difficulty perceiving f_0 information using temporal envelope cues for both 100 and 200 Hz reference stimuli, and therefore it could be expected that the perception of spectrally degraded male and female voices would be equally affected by aging. Overall, these results confirm previous findings that used neurophysiologic and electrophysiologic methods, and consequently lend further support for senescent decline in temporal envelope pitch processing.

For example, neurophysiologic experiments in laboratory animals show that single-unit responses to SAM noise recorded from the inferior colliculus (IC) also change as a function of age (Walton *et al.*, 2002). Walton *et al.* (2002) recorded single-unit responses in the IC to SAM noise in young and old CBA mice. Stimuli were 100% amplitude modulated, and modulation frequencies varied from 10 to 800 Hz. Results showed an increase in spike count and phase-locking among older mice. Units from older mice tended to produce greater spike counts and phase locking to multiple SAM rates, whereas units from younger mice tended to exhibit lower spike counts and better rate specificity. The strongest responses among younger mice occurred at higher modulation frequencies (200–400 Hz), while the strongest responses among older mice occurred at lower modulation frequencies (40–100 Hz). However, in the present study, we found that f_0 discrimination using temporal envelope cues (one-channel condition) did not depend on the reference f_0 . We did not find that older listeners' performance was worse with a higher modulation rate (200 Hz) compared to a lower modulation rate (100 Hz) for any of the NBV conditions, as might be expected given the results of previous studies (Grose *et al.*, 2009; Purcell *et al.*, 2004; Walton *et al.*, 2002).

The lack of significant interactions between age group and reference f_0 is somewhat surprising based on these previous studies. However, two separate issues might account for the difference between the results reported in the present study and those of previous studies. The first is related to the specific modulation frequencies and/or f_0 s tested. For example, Grose *et al.* (2009) found an interaction between aging and modulation frequency when they measured Auditory Steady State Responses (ASSR) in normal hearing listeners. Modulation frequencies used to measure the ASSR were 32 and 128 Hz. In fact, studies measuring ASSR in older adults show no age effect for very low frequencies (<40 Hz) but do find an age effect in higher frequencies (>128 Hz). As discussed previously, Walton *et al.* (2002) found that strongest responses among younger mice occurred at higher modulation frequencies (200–400 Hz), while the strongest responses among older mice occurred at lower modulation frequencies (40–100 Hz). Therefore, age-related differences in periodicity coding may be dependent on the frequency of the stimulus within a broader range of modulation frequencies than those used in the present experiment.

One must also take into consideration the unit choice when analyzing results. The current study used log transformed data for all analyses, while it appears that some previous studies chose to analyze data using a linear scale (for example, Grose *et al.*, 2009, among others). This decision will obviously affect the relative magnitudes of the age effects observed at each reference frequency and could influence the overall outcome of the results. In fact, when data in the current study are analyzed using a linear scale, it changes the interpretation of the data significantly; a significant interaction is noted between age group, spectral degradation, and reference f_0 . Specifically, between-group differences in f_0 discrimination increase as the number of spectral channels is reduced. This effect is greater for the 200 Hz reference condition. The Shapiro-Wilk test was performed on both the linear (raw) data set as well as the log transformed data set. We found that normality was violated for all conditions when data were analyzed in linear form; however, normality was not violated when data were analyzed in log units. Because the ANOVA assumes a normal distribution of data, the data can be more accurately interpreted when converted to log units.

The results of the current study are generally in keeping with those of Schwartz and Chatterjee (2012) in which older listeners exhibited poorer performance than younger listeners when measured on a gender identification task using stimuli of varying spectral and temporal cues. However, those results suggested that older listeners show particular difficulty discriminating voice-pitch when they are forced to rely on temporal envelope cues to do so (one-channel condition). This finding is in contrast with the current study which showed overall senescent decline in periodicity coding, for both spectro-temporal and temporal-envelope based perception. This interpretation and comparison, of course, assumes that in the current study subjects were able to use some spectral cues as the number of channels increased. On the basis of the results, this assumption seems valid but nonetheless there is no way to specifically know which cues listeners

were attending to for a given condition. Several cues other than f_0 also contribute to the total perception of voice gender (Klatt and Klatt, 1990; Murry and Singh, 1980). Therefore, it was important to isolate effects of aging on use of the f_0 cue alone. The current findings are also in agreement with those reported by Souza *et al.* (2011) who found that f_0 discrimination using an eight channel NBV vowel stimulus was also age dependent.

Research suggests that noise carriers can create increased modulation interference (Dau *et al.*, 1999). Specifically, the intrinsic fluctuations of the noise carrier can interfere with detection of an imposed temporal envelope modulation; the degree of interference depends on the noise type and noise bandwidth. Such interference might have well been present in our experiments, and might have reduced listeners' ability to process temporal envelope cues overall. The fact that older listeners did not show significant differences in the overall pattern than younger listeners, suggests that they are not in fact more susceptible to the effects of modulation interference than younger listeners, at least in the context of our experimental tasks.

Outcomes of the current investigation suggest that older listeners' reduced ability to discern voice gender when spectral cues are reduced, are at least in some part due to impaired periodicity coding. While the current study suggests older listeners may have difficulty using temporal envelope pitch information when stimuli are presented in isolation, further studies are needed to determine how older listeners might utilize f_0 cues to perceptually segregate simultaneous spectrally degraded stimuli (e.g., speech understanding in the presence of another talker) which would be more indicative of real-world listening. Regardless, results of the current study could have important implications for the counseling and rehabilitation of older CI recipients regarding expectations and performance.

ACKNOWLEDGMENTS

We thank Dharma A. Teja for the development of experimental software, and to the individuals who participated in the study. This project was funded by National Institutes of Health/National Institute on Deafness and Other Communication Disorders Grant No. R01DC004786 (MC), and training Grant No. T32DC000046 awarded to K.C.S.-L. (PI: Arthur Popper).

Anderson, S., Parbery-Clark, A., Yi, H.-G., and Kraus, N. (2011). "A neural basis of speech-in-noise perception in older adults," *Ear Hear* **32**, 750–757.

ANSI (2004). S3.6-2004, *American National Standard Specification for Audiometers* (American National Standards Institute, New York).

Arehart, K. H. (1994). "Effects of harmonic content on complex tone fundamental frequency discrimination in hearing-impaired subjects," *J. Acoust. Soc. Am.* **95**, 3574–3585.

Bernstein, J. G., and Oxenham, A. (2003). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?," *J. Acoust. Soc. Am.* **113**, 3323–3334.

Bernstein, J. G., and Oxenham, A. (2006). "The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss," *J. Acoust. Soc. Am.* **120**, 3929–3945.

Broxk, J. P. L., and Nootebohm, S. G. (1982). "Intonation and the perceptual separation of simultaneous voices," *J. Phonetics* **10**, 23–36.

Brown, C. A., and Bacon, S. P. (2010). "Fundamental frequency and speech intelligibility in background noise," *Hear. Res.* **266**, 52–59.

Brungart, D. S. (2001). "Informational and energetic masking effects in the perception of two simultaneous talkers," *J. Acoust. Soc. Am.* **109**, 1101–1109.

Chao, Y. R. (1968). *A Grammar of Spoken Chinese* (University of California Press, Berkeley, CA), 847 pp.

Chatterjee, M., and Oberzut, C. (2011). "Detection and rate discrimination of amplitude modulation in electrical hearing," *J. Acoust. Soc. Am.* **130**, 1567–1580.

Chatterjee, M., and Peng, S. C. (2008). "Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition," *Hear. Res.* **143**, 156–235.

Dau, T., Verhey, J., and Kohlrausch, A. (1996). "Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers," *J. Acoust. Soc. Am.* **99**, 3615–3622.

Drolet, M., Schubotz, R. I., and Fischer, J. (2014). "Recognizing the authenticity of emotional expressions: F0 contour matters when you need to know," *Front. Human Neurosci.* **8**(144), 1–11.

Fishman, K. E., Shannon, R. V., and Slattery, W. H. (1997). "Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor," *J. Speech Language Hear. Res.* **40**, 1201–1215.

Folstein, M., Folstein, S. E., and McHugh, P. R. (1975). "'Mini-Mental State' a practical method for grading the cognitive state of patients for the clinician," *J. Psychiatric Res.* **12**, 189–198.

Fornby, C. (1985). "Differential sensitivity to tonal frequency and to the rate of amplitude modulation of broadband noise by normally hearing listeners," *J. Acoust. Soc. Am.* **78**, 70–77.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (2001). "Speech recognition in noise as the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.* **110**, 1150–1163.

Fuller, C. D., Gaudrain, E., Clarke, J. N., Galvin, J. J., Fu, Q.-J., Free, R., and Başkent, D. (2014). "Gender categorization is abnormal in cochlear implant users," *J. Assoc. Res. Otolaryngol.* **15**, 1037–1048.

Galvin, J. J., Oba, S., Baskent, D., and Fu, Q.-J. (2015). "Modulation frequency discrimination with single and multiple channels in cochlear implant users," *Hear. Res.* **324**, 7–18.

Grant, K. W., Summers, V., and Leek, M. R. (1998). "Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **104**, 1051–1060.

Greenwood, D. D. (1990). "A cochlear frequency-position function for several species—29 years later," *J. Acoust. Soc. Am.* **87**, 2592–2605.

Grose, J. H., Mamo, S. K., and Hall, J. W. (2009). "Age effects in temporal envelope processing: Speech unmasking and auditory steady state responses," *Ear Hear.* **30**, 568–575.

Klatt, D. H., and Klatt, L. C. (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Am.* **87**, 820–857.

Kong, Y.-Y., Stickney, G. S., and Zeng, F.-G. (2005). "Speech and melody recognition in binaurally combined acoustic and electric hearing," *J. Acoust. Soc. Am.* **117**, 1351–1361.

Lehiste, I. (1970). "Suprasegmental features of speech," in *Contemporary Issues in Experimental Phonetics*, edited by N. J. Lass (Academic Press, New York), pp. 225–239.

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.

Luo, X., Fu, Q. J., Wei, C. G., and Cao, K. L. (2008). "Speech recognition and temporal amplitude modulation processing by Mandarin-speaking cochlear implant users," *Ear Hear.* **29**, 957–970.

Moore, B. C. J., and Peters, R. W. (1992). "Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity," *J. Acoust. Soc. Am.* **91**, 2881–2893.

Murry, T., and Singh, S. (1980). "Multidimensional analysis of male and female voices," *J. Acoust. Soc. Am.* **68**, 1294–1300.

Purcell, D. W., John, S. M., Schneider, B. A., and Picton, T. W. (2004). "Human temporal auditory acuity as assessed by envelope following responses," *J. Acoust. Soc. Am.* **116**, 3581–3593.

Qin, M. K., and Oxenham, A. J. (2005). "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," *Ear Hear.* **26**, 451–460.

Schwartz, K. C., and Chatterjee, M. (2012). "Gender identification in younger and older adults: Use of spectral and temporal cues in noise-vocoded speech," *Ear Hear.* **33**, 411–420.

- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **279**, 303–304.
- Smalt, C. J., Krishnan, A., Bidelman, G. M., Ananthakrishnan, S., and Gandour, J. T. (2012). "Distortion products and their influence on representation of pitch-relevant information in the human brainstem for unresolved harmonic complex tones," *Hear. Res.* **292**, 26–34.
- Souza, P., Arehart, K., Miller, C. W., and Muralimanohar, R. K. (2011). "Effects of age on F0-discrimination and intonation perception in stimulated electric and electro-acoustic hearing," *Ear Hear.* **32**, 75–83.
- Takahashi, G. A., and Bacon, S. P. (1992). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," *J. Speech Hear. Res.* **35**, 1410–1421.
- Titze, I. R. (1989). "Physiologic and acoustic differences between male and female voices," *J. Acoust. Soc. Am.* **85**, 1699–1707.
- Vongpaisal, T., and Pichora-Fuller, M. K. (2007). "Effect of age on F0 difference limen and concurrent vowel identification," *J. Speech Language Hear. Res.* **50**, 1139–1156.
- Walton, J. P., Simon, H., and Frisina, R. D. (2002). "Age-related alterations in the neural coding of envelope periodicities," *J. Neurophysiol.* **88**, 565–578.