HEARING ONE OR TWO VOICES : F0 AND VOWEL SEGREGATION IN YOUNGER AND OLDER ADULTS

Tara Vongpaisal and Kathy Pichora-Fuller
Dept. of Psychology, University of Toronto at Mississauga, 3359 Mississauga Rd. N., Ontario, Canada, L5L 1C6.
tarav@psych.utoronto.ca; kpfuller@utm.utoronto.ca

In the present study, we investigate the effect of age on the ability to detect differences in the fundamental frequency (F0) of vowels and the use of this cue to identify simultaneous vowels.

Older listeners with good hearing in the speech range experience relatively little difficulty understanding one talker in a quiet listening environment. However, they report difficulty understanding speech in multi-talker situations. Age-related changes in auditory temporal processing are believed to contribute to these difficulties (Schneider & Pichora-Fuller, 2001; Pichora-Fuller & Souza, 2003). Periodicity coding is one aspect of auditory temporal processing. It enables a listener to use information conveyed by the F0 and harmonic structure of speech. This information is important for the perception of voice pitch and quality (Assmann & Summerfield, 1990). Voice pitch and quality may help listeners to segregate voices in multi-talker situations.

Compared to normal hearing listeners, hearing impaired listeners are less able to detect differences in F0—a problem that is associated with the reduced ability to identify concurrently spoken vowels (e.g., Summers & Leek, 1998). Although Summers and Leek (1998) set out to examine the effects of hearing loss on F0 detection and concurrent vowel identification, they noticed that regardless of audiometric status, older listeners performed more poorly than younger listeners. In the present study, the effect of age on F0 detection and its relationship to vowel identification is directly investigated in listeners with good audiograms.

Experiment 1

Method

Participants. Fifteen younger adults (M = 26 years of age, SD = 3.2) and 15 older adults (M = 74 years of age, SD = 5.6) were recruited from the Mississauga area. All had good hearing with pure-tone, air-conduction thresholds ≤ 25 dB HL between .25 and 3 kHz in their better ear. These participants also completed Experiment 2.

Stimuli and Apparatus. Tokens of the vowel [a] were synthesized using five fixed formant frequencies (Assmann and Summerfield, 1994) and with F0 varying from 120 to 130Hz in increments of .1 Hz. The synthesized vowel stimuli were presented through a TDT System II and presented monaurally through TDH49 headphones at 80 dB SPL. Testing took place in a double-wall sound-attenuating booth (Industrial Acoustics Corporation).

Procedure. A practice block of 100 trials was administered before the test phase. Each trial of the practice consisted of three successively presented tokens (duration = 260 ms) of the vowel [a]: the standard token (F0 = 120 Hz) followed by two comparisons tokens. The interstimulus interval was 150 ms. One of the comparison tokens matched the standard while the other differed from the standard (F0 = 145 Hz). Listeners indicated which of the two comparison tokens was different from the standard by pressing the corresponding button on a button box. Feedback was given after each trial. In the test phase, the F0 difference limen threshold for each participant was determined using an adaptive task. The initial step size was 30 Hz and the step size on subsequent trials was halved following a correct response or doubled following three incorrect responses. After five reversals, the increments were reduced from 2 to 1.25 and decrements were increased from .5 to .8. F0 was determined from the mean of the last 10 reversals. Each participant completed the adaptive task three times. The final ΔF0 threshold was the average of the three runs.

Results and Discussion

In the practice block, both groups achieved a high degree of accuracy in detecting the contrast from the standard token: younger adults (M = 98.1%, SD = 2.3), older adults (M = 94.3%, SD = 7.8). Younger adults had reliably lower ΔF0 thresholds (M = .6 Hz, SD = .4, range = .1 to 1.0 Hz) than older adults (M = 1.8, SD = .8, range = .3 to 3.1), $t(28)$ = 5.063, $p < .001$. Not surprisingly, the F0 thresholds for the younger adults were better than those found previously for middle-aged adults, while the F0 thresholds for older adults were similar to those found previously for older adults with normal hearing and better than those of older adults with hearing loss (Summers & Leek, 1998). The reduced ability of older adults to detect differences in F0 could reduce their ability to use this cue in segregating voices.

Experiment 2

Method

Apparatus and Stimuli. Five synthesized vowels [a, i, æ, er, u] were created using formant frequencies corresponding to Assmann and Summerfield (1994). There were six tokens of each vowel differing in F0 (F0 = 120, 122, 124, 127, 135, 151 Hz, respectively corresponding to increases of 0, .25, .5,

1, 2, and 4 semitones from an F0 of 120 Hz). All possible paired combinations of vowels were formed. Thus, a total of 150 vowel pairs were created (5 x 5 pairs x 6 F0 difference levels). All stimuli were 260 ms in duration and delivered monaurally to the better ear via headphones at an overall level of 80 dB SPL.

Procedure. Listeners were first familiarized with all 150 vowel tokens in a single vowel identification task. They identified each vowel from a closed set of five vowel alternatives. Feedback was provided after each response and this procedure was repeated as necessary until listeners achieved an accuracy of at least 80%. Next, listeners completed the concurrent vowel labeling task. They heard pairs of simultaneously presented vowels and identified each vowel by pressing the corresponding vowel button on the button box. If listeners were uncertain on the identity of the vowel, they were instructed to select their best guess. No feedback was provided.

Results and Discussion

All participants identified single vowels with a high degree of accuracy, with younger adults obtaining a higher proportion of correct responses (M = .98, SD = .02) than older adults (M = .93, SD = .06), $t(28) = 3.209$, $p = .003$.

Figure 1 shows the mean proportion of correct identification of both vowels by the younger and older groups as a function of F0 difference in the concurrent vowel labeling task. Overall, a mixed design analysis of variance (2 age groups x 6 F0 difference levels) revealed a significant main effect for age group, with younger adults performing with greater accuracy (M = .58 , SD = .09) than older adults (M = .31, SD = .04) across all F0 difference levels, $F(1, 28) = 21.06$, $p < .001$. There is also a significant main effect for F0 difference, $F(5, 140) = 30.56$, $p < .01$, as well as a significant two-way interaction, $F(5, 140) = 4.42$, $p = .001$. To examine the interaction, Tukey HSD tests were conducted to determine whether each group's performance differed across F0 difference levels. For young adults, there was a significant increase ($p < .001$) in accuracy when F0 separation increased from 0 to .25 semitones. However, for older adults, the improvement in accuracy when F0 separation increased from 0 to .25 semitones approached, but did not reach, conventional levels of significance (p = .068). Beyond an F0 difference of .25 semitones, there was no significant increase in accuracy for younger and older adults ($ps > .05$). Subsequent analyses on the individual data of older adults reveal that three of these participants performed within 1 SD of the mean of younger adults across all F0 difference levels. All others performed significantly below younger adults, but above chance performance.

Simple correlations were conducted to examine associations between participants' performance on the smallest F0 separation values (i.e., 0, and .25 semitones) with age and ΔF0 threshold. Both variables correlated negatively with performance at zero semitone separation: age, $r(29) = -.669$, $p < .001$; and ΔF0 threshold, $r(29) = -.563$, $p = .001$. Similarly, performance at .25 semitone separation is negatively correlated with age, $r(29) = -.651$, $p < .001$, and with ΔF0 threshold, $r(29) = -.594$, $p = .001$.

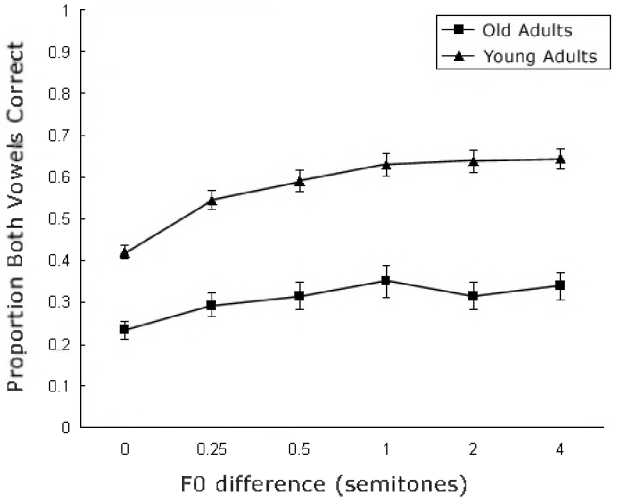

Fig. 1. Proportion of correct identification of both vowels as a function of F0 difference. Error bars indicate standard error.

Analyses of partial correlations revealed that both age and ΔF0 thresholds retained reliable associations with performance at both of these F0 separation values. In sum, these correlations suggest that both age and ΔF0 threshold have a role in the ability to segregate and identify simultaneously presented vowel sounds.

In summary, there are age differences in the ability to detect ΔF0 and the ability to use this cue to identify concurrent vowels. These findings support the hypothesis of loss of periodicity coding as a characteristic of auditory aging (Schneider & Pichora-Fuller, 2001). It sheds new light on older listeners' difficulties in using voice pitch and quality to follow speech in multi-talker environments.

References

Assmann, P. F., & Summerfield, Q. (1990). Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. *Journal of the Acoustical Society of America, 88,* 680-697.

Assmann, P. F., & Summerfield, Q. (1994). The contribution of waveform interactions to the perception of concurrent vowels. *Journal of the Acoustical Society of America, 95,* 471-484.

Pichora-Fuller, M. K., & Souza, P. E. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology, 42,* 2S11-2S32.

Schneider, B., & Pichora-Fuller, M. K. (2001). Age-related changes in temporal processing: Implications for speech perception. *Seminars in hearing, 22,* 227-239.

Summers, V, & Leek, M. R. (1998). F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *Journal of Speech, Language, and Hearing Research, 41,* 1294-1306.