Effects of Aging on Vocal Fundamental Frequency and Vowel Formants in Men and Women

Julie Traub Eichhorn, Raymond D. Kent, Diane Austin, and Houri K. Vorperian, *Madison, Wisconsin*

Summary: Purpose. This study reports data on vocal fundamental frequency (f_o) and the first four formant frequencies (F1, F2, F3, F4) for four vowels produced by speakers in three adult age cohorts, in a test of the null hypothesis that there are no age-related changes in these variables. Participants were 43 men and 53 women between the ages of 20 and 92 years.

Results. The most consistent age-related effect was a decrease in f_o for women. Significant differences in F1, F2, and F3 were vowel-specific for both sexes. No significant differences were observed for the highest formant F4.

Conclusions. Women experience a significant decrease in f_o, which is likely related to menopause. Formant frequencies of the corner vowels change little across several decades of adult life, either because physiological aging has small effects on these variables or because individuals compensate for age-related changes in anatomy and physiology.

Key Words: Adult acoustics–Aging voice–Fundamental frequency–Formants–Sex differences.

INTRODUCTION

From infancy through young adulthood, the acoustic signal of speech undergoes substantial change, including marked decreases in both vocal fundamental frequency (f_o) and the frequencies of the vowel formants.[1] These developmental changes are accompanied by a sexual dimorphism that is proportionately (ie, male-to-female ratio) one of the largest in human development.[2] After adulthood, age-related changes in the acoustic properties of speech and voice are much less marked and conclusions across studies are inconsistent, with some studies showing no effects and others reporting a variety of effects such as a decrease of f_o in women,[3,4] an increase of f_o in men,[3,4] a centralization of formant frequencies,[5] a decrease in F1 frequency,[6] decreases in all formant frequencies,[7] and a sex-vowel interaction in formant frequencies.[8]

Because results are not consistent across reports and because the aggregate number of individuals studied is small compared to the general population, it is difficult to draw firm conclusions on the effect of aging on speech acoustics. Even if acoustic changes occur with age, there are uncertainties in the interpretation of age- and sex-related variations. For example, changes in vowel formant frequencies during adulthood have been linked to lengthening of the vocal tract,[7,9] altered dimensions of the back cavity of the vocal tract,[10] diachronic or intergenerational phonetic change,[11,12] reduction of articulatory movement,[13] and adjustments of lingual articulation.[14] Possibly, two or more of these factors operate in combination to account for age-related acoustic changes in speech, and the combinations may vary among individuals.

Given the diverse results in previous studies, it is difficult at this time to describe a normative lifespan pattern for measures of f_o and formant frequencies in vowel production. A normative pattern is needed to inform studies of quality of life during aging,[15] to serve as reference data for clinical assessment and treatment,[16–18] to provide information for biometric identification and forensics[19] and speech technologies such as automatic speaker recognition.[20] Most of the relevant research to date has focused on f_o and other features of phonation. A much smaller literature has been published on the joint effects of aging on f_o and formant frequencies, and most of these reports present data only for F1 and F2, neglecting the higher formants, which may be sensitive to sex differences and alterations in vocal tract geometry.

More generally, little is known about the effects of aging on any aspect of speech production. Smiljanic[21] concluded that, "In contrast to the accumulated knowledge about the perceptual processing difficulties [in aging], very little is known about whether age-related changes impact speech production patterns for older adult talkers and the intelligibility of their speech" (p. EL129). Research to date, although limited, indicates that older adults have reduced speech intelligibility,[22] greater impairment of phonological than semantic levels of language production,[23] and difficulties with specific articulatory features.[24] Acoustic studies shed light on why intelligibility changes with age. For example, Benjamin[25] concluded that aging affected vowel productions, voice onset time, phoneme segment duration, and speaking rate. Schötz[26] reported age-related changes in speaking rate (segment duration), intensity range, and, to a lesser extent, f_o and the frequencies of the first two formants. She also noted that the acoustic correlates of aging speech are not the same in men and women. Because multiple acoustic cues underlie speech intelligibility, systematic research is needed to determine which cues, perhaps in various combinations, explain reduced intelligibility. Changes that occur in healthy aging are an important basis for understanding speech and voice disorders associated with health conditions that affect older individuals, such as hearing loss, Parkinson's disease, dementia, stroke, and cancer.[27]

The purpose of this study is to test the null hypothesis that there are no age-related changes in f_o and the first four formant frequencies (F1, F2, F3, F4) for four vowels produced by speakers in three adult age cohorts. Several features of this research are notable. The test words used in this report are from a larger lifespan study of speech acoustics that covers the age range of

Accepted for publication August 1, 2017.
From the Waisman Center, University of Wisconsin-Madison, Madison, Wisconsin.
Address correspondence and reprint requests to Houri K. Vorperian, 427 Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, WI 53705. E-mail: vorperian@waisman.wisc.edu; hkvorper@wisc.edu
Journal of Voice, Vol. ■■, No. ■■, pp. ■■-■■
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2017.08.003

4 to 92 years. Constancy of the words across speakers throughout the life span facilitates comparisons and reduces variability related to the use of different speech samples across speaker ages. Collection of data for the first four formants, as opposed to just two formants, as has been the case in most studies, provides additional information that may be helpful in explaining age-related changes in formant pattern, for example, by referring to formant-cavity affiliations.

METHODS

Participants

Speech recordings were from 96 healthy adult participants (43 male, 53 female), between the ages of 20 and 92 years. All participants met eligibility requirements as native English speakers. Ages were grouped into three cohorts: cohort I consisted of young adults (19 males, 21 females) ages 20–30 years, cohort II consisted of middle-aged adults (12 males, 20 females) ages 40–60 years, and cohort III consisted of older adults (12 males, 12 females) ages 70–92 years. No participant was excluded on the basis of hearing status. Not surprisingly, the majority of participants in cohort III had a hearing loss as determined by self-report or a screening test (nine wore hearing aids).

Speech sample

The speech stimuli consisted of 20 unique monosyllabic American English words. The words had the syllable structures of consonant-vowel, vowel-consonant, or consonant-vowel-consonant. The words were composed with the four corner vowels in the classic vowel quadrilateral: /i/ (bead, bee, eat, sheep, feet), /u/ (boo, boot, zoo, hoot, shoe), /æ/ (bath, bat, cat, hat, sad), and /ɑ/ (dot, hop, pot, top, hot). For each vowel, two of the stimuli were recorded twice (eg, bead, eat, bat, hat). The stimuli were designed to collect data over the life span and were therefore chosen to be familiar to young children, and also to have high phonological neighborhood density, which reportedly maximizes F1/F2 acoustic vowel space.[28]

Recording protocol

Participants were recorded in a quiet room in either a laboratory or a retirement center. Background noise levels were measured with a Fisher Scientific Sound Level Meter Model 11-661-6A (Distributed by Control Company, Friendswood, TX; manufactured in Taiwan) with an A-weighting. The levels varied between 31 and 38 dBA, depending on the recording site. Recordings were made with a Shure-SM48 (Manufactured by Shure, Juarez, Mexico) microphone mounted on a floor stand and attached to a Marantz digital recorder (Distributed by Martel Electronics, Inc. Yorba Linda, CA. Manufactured in Japan). The microphone was adjusted to each participant's seated height and positioned at an approximately 15-cm distance from the mouth. The Marantz-PMD660 digital audio recorder digitizes speech at 48 kHz with 16-bit resolution on a SanDisk Ultra II flash-card (Manufactured in China). To optimize recording level, the Marantz recorder gain was adjusted to 6–12 dB below the maximum level on the volume unit meter. Using a laptop with the *TOCS+ Platform program*[29] for randomization, the stimuli were presented visually and aurally using pictures with the orthographic word, while playing the recording of an adult male through external speakers. Participants were instructed to repeat the words at a normal loudness level. Two practice words were used at the beginning to adjust recording levels. During recording, stimuli that were mispronounced or considered to be deviant in volume unit meter reading were repeated and the repeated recording replaced the original productions.

Acoustic analysis methods

Acoustic analyses were based on methods and criteria developed for the analysis of speech from speakers who represent various combinations of age and sex.[30,31] The basic steps were as follows: speech recordings were uploaded to a computer, segmented into separate word files with *Praat* (version 5.1.31 [Computer Software] Amsterdam, The Netherlands by Boersma and Weenink),[32] and saved into separate wave files. The vowel portion of each word was analyzed to obtain estimates of f_o and the first four formants (F1, F2, F3, F4) using the acoustic analysis software *TF32* (Milenkovic 2010 Madison, WI, USA).[33] The frequency of f_o was measured with the pitch determination algorithm in *TF32*. The formant frequencies were estimated with reference to the spectrogram (with overlaid formant tracks determined by linear prediction coding [LPC]) and time-slice spectra from LPC and fast Fourier transform (FFT). The LPC and FFT spectra were superimposed to permit comparisons. In cases where the LPC and FFT spectra were not congruent, reference was made to the spectrogram together with adjustments in analysis parameters, especially the number of LPC coefficients. The default setting for *TF32* is a 300 Hz bandwidth and 48 dB dynamic range. Both the number of LPC coefficients and the dynamic range were adjusted to achieve optimum results for each speaker. Males were analyzed at 300 Hz, with a dynamic range adjusted between 48 and 64 dB. Females were analyzed at default settings; however, if formant energy was unclear, bandwidth was adjusted to 350 or 400 Hz and dynamic range could be adjusted as high as 68 dB. These procedures are consistent with general practice and recommendations for formant analysis using LPC and FFT.[34–37] Because we have observed that the temporal midpoint of the vowel does not always correspond to a formant steady state, we used vowel-specific time points as defined by Derdemezis et al[31] and similar to criteria used by others:[38]

(a) Vowel /i/—point of highest F2 frequency;
(b) Vowel /u/—point of lowest F2 frequency;
(c) Vowel /ɑ /—point of least separation of F1 and F2 frequencies;
(d) Vowel /æ/—point of most evenly spaced formants, taking care to avoid measurement at a point of decreasing F2-F1 difference which can reflect backing of the vowel.

The value of f_o was determined at the same time point as the formant measurements unless this time point occurred during an interval of vocal fry, pitch break, or other deviation from the overall f_o contour of the syllable. In such cases, f_o was measured at a proximal time point where the value was consistent