LONGITUDINAL INVESTIGATION OF THE AGING VOICE

Benjamin V. Tucker[1,2], Stephanie Hedges[2], Mark Berardi[3], Eric J. Hunter[4],

[1]Northern Arizona University, [2]University of Alberta, [3]University Hospital Bonn, [4]Michigan State University
benjamin.tucker@nau.edu, shedges@ualberta.ca, Mark.Berardi@ukbonn.de, ejhunter@msu.edu

ABSTRACT

Acoustic and perceptual research has shown that an individual's vocal characteristics change with age. In this paper, we report a later-life longitudinal investigation of three male speakers using publicly available archives of speeches given to large audiences on a semi-regular basis (generally with a couple of years between each instance). The collection of speeches was given during the last 40-50 years of each speaker's life. From each speech, 5-minute samples (recordings and transcripts) were force-aligned to identify word and phoneme boundaries. We investigate non-linear changes in vowel duration, word duration, speech rate, fundamental frequency, and vowel space using Generalized Additive Mixed Models. We describe the vocal changes over the lifespan in the speakers' later years. The current results do not support expected outcomes from previous research and illustrate idiosyncratic and dynamic effects for individuals in later years.

Keywords: speech production, aging, duration, fundamental frequency, vowel space

1. INTRODUCTION

With advancing age, our bodies undergo a variety of physiological and cognitive changes that impact speech [1]. These changes have been shown to impact speech production and speech acoustics metrics. In order to understand how age influences speech, we use longitudinal data from three separate male speakers covering the last 40-50 years of life. This paper investigates the longitudinal effects of age on vowel and word duration, speech rate, vowel space, and fundamental frequency.

Research on speech and aging has found variable effects [2, for a review]. Generally, segment and word duration increases with advancing age [3, 4, 5]. Jacewisz et al. [6] found that reading and speaking rate increased from early adulthood to middle adulthood following which it slowly decreased throughout late adulthood. However, Quene et al., [7] found that the speech rate of one Dutch speaker (Queen Beatrix) between the ages of 42 and 74 years decreased from 42 to 62 years of age and then increased from 62 to 74 years. They argue that it is possible that the increased speech rate in Dutch culture and decades of experience reading the *Troonsrede* may have impacted her speech rate.

Many studies have found changes in fundamental frequency with age [8, 9, 10, 11, 12, 13, 14]. Researchers performing long-term longitudinal investigations have primarily focused on individual speakers' fundamental frequency changes. For example, Harrington found a significant decrease of an average of 21 percent in fundamental frequency for four speakers from an earlier recording to a recording made at a later age [10]. In a more in-depth analysis of the speech of both Queen Elizabeth II and Alistair Cooke, Reubold et al. [11] found that the fundamental frequency of Queen Elizabeth decreased as a function of age in a linear pattern while for Alistair Cooke fundamental frequency decreased until the age of 87, but then increased after 87. Hunter et al. [15] used one of the same subjects as the current study (S1) and found that F0 decreased until the age of 68 where it then increased until the age of 98 years. They attribute this change to general physiological changes. In cross-sectional studies, researchers have found differences in the patterns of male and female participants with a u-shaped result for female participants [13, 14, 3].

The vowel space of aging speakers has also been investigated. This research is also variable with some research indicating an increase in phonetic height (decreased F1) and backness (decreased F2) with age [10, 11, 1, 16]. In a longitudinal analysis of the speech of participants from the UP corpus, Gahl & Baayen [17] examined how the vowel space shifts with aging between early to late adulthood. They found that vowel space shifts more toward the periphery with an increase of age, even when controlling for an increase in vowel duration.

The purpose of this study was to build on previous reports of later-life speech changes. From the perspective of using available recordings and automatic analysis techniques, later-life speech changes could be investigated over a larger period of time. This approach allows for a larger number of longitudinal samples compared to what could be accomplished in a laboratory setting.

2. METHOD

The present study analyzes regular speeches by three religious leaders to university students that cover a time period ranging from 40 to 50 years. This source includes longitudinal data covering middle adulthood to late adulthood. The speeches were given in the same style keeping the style characteristics consistent across the years recorded.

2.1. Participants

This study analyzes the speech of three male speakers: S1 (born: 1910), S2 (1899), and S3 (1922). Speakers S1 and S3 were born in Utah, and S2 was born in Idaho. The majority of the speakers' lives were spent in the western United States, especially Utah and Idaho. However, both S2 and S3 spent time in the eastern United States living and working in Washington DC, Massachusetts, and New York. Additionally, per lifestyle choice, the speakers reportedly did not consume alcohol or coffee, and they did not smoke; these substances all have been shown to affect voice production and vocal health [18, 19].

2.2. Forced Alignment

Five-minute segments were extracted from each recorded speech and were then annotated orthographically using the linguistic annotator ELAN software [20]. Phonetic alignment of the speech was obtained using the Penn Forced Aligner [21] with no manual correction. We extracted acoustic measures using Praat [22] for 73,305 stressed and unstressed vowels from 51,842 words.

3. RESULTS

We performed a generalized additive mixed models (GAMM) [23, 24] on the extracted data to account for non-linear trends found within the acoustic trajectories over time with speaker and word included as random effects. Four separate GAMMs were run on the data to investigate trends for vowel duration, word duration, speech rate (syllables per second for each recording), and fundamental frequency with advancing `age` as a smoothed predictor and word and speaker as random effects. Convex hulls for each speaker's vowel space were calculated in order to investigate vowel space shifts

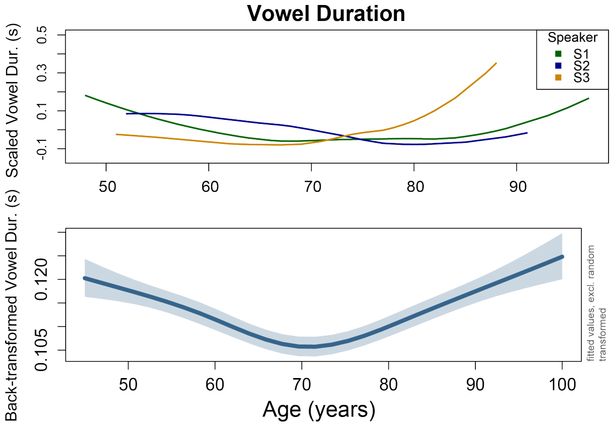

Figure 1: Mean vowel duration overtime for S1, S2, and S3

with age [25, 26].

Figure 1 shows the mean vowel duration of the stressed vowels ($n = 41,359$) for each speaker (top) and a visualization of the GAMM results (bottom). Each speaker shows a unique pattern of mean vowel duration over the later-decades of the lifespan. Vowel duration for S1 decreases to about the age of 68 and then vowel duration starts to increase at the age of 85. While vowel duration for S2 generally decreases to about the age of 80 with a slight increase after that. S3, interestingly, shows a largely flat, with maybe a slight decrease to about the age of 70 followed by a sharp increase in duration, producing the longest vowel durations of all the speakers. The GAMM with duration measures for all three speakers, illustrated in lower part Figure 1, shows that the modeled vowel duration trajectory follows a U-shaped pattern, with vowel duration decreasing until approximately the age of 71 years, where it then increases. The model shows a significant effect of `Age` ($edf = 3.8144, F = 66.4155, p < 0.0001$).

The average word duration data by speaker is plotted at the top of Figure 2 and the effect of `Age` in the GAMM model is found at the bottom of Figure 2. The top part of the figure illustrates the individual variation present for each speaker. While the word duration begins at about the same point in their early 50s, S1 decreases slightly, then increases slightly where word duration then levels off for the last decade and a half. For S2, however, there is a slight increase after which there is a U-shaped pattern of a decrease followed by an increase occurs. The trajectory for S3 shows a relatively stable duration until his mid-70s, where it then sharply increases. The bottom part of Figure 2 summarizes the result of a GAMM with Age as the primary predictor of word duration for all speakers. The model indicates that as speakers age word duration follows a U-shaped pattern, similar to that which was found for vowel duration. From the ages of approximately 48 to 70 years, word duration is flat or decreases. Following the age of 70 years, words are produced with longer durations as speakers age ($edf = 3.9258, F = 81.3261, p < 0.0001$).

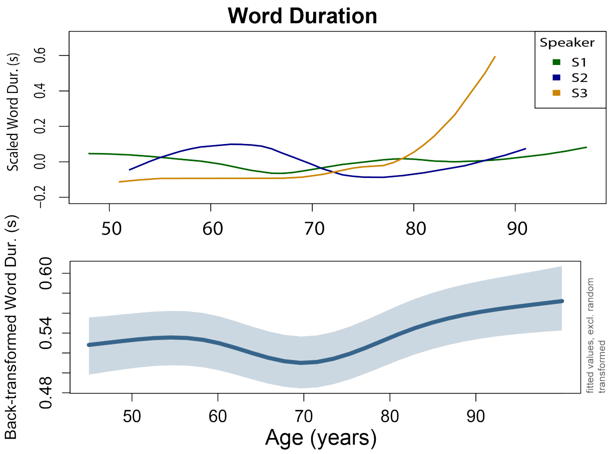

Figure 2: Word duration over time for S1, S2, and S3 and the results of a GAMM analysis of word duration (s) over time.

The average speech rate for each speaker (the number of vowels divided by the total duration of the recording) is plotted in the top half of Figure 3. As was the case with both vowel and word duration, the average speech rate pattern is unique to each individual. With S1, speech rate increased until around the age of 68 years where it then seemed to level off until the age of approximately 81 years, where it then decreased. S2 followed a somewhat similar trajectory with an increase of speech rate until the age of around 80 years, followed by a decrease. The speech rate of S3 increased until the age of 68 years, when rate decreases sharply. The pattern for all three speakers is generally consistent; however, the age at which the decrease in speech rate begins varies across the three speakers. We generated a GAMM for speech rate using data from all three speakers again focusing on the effect of Age. In the model, the trajectory of speech rate was found to increase until the age of 70 years, where it then decreases ($edf = 2.983, F = 2.84, p < 0.05$). This aligns nicely with the results reported for both vowel and word duration.

The top part of Figure 4 plots the average fundamental frequency of the stressed vowels for each speaker as they age. While a general U-shaped pattern can be seen for each speaker, there are differences between the speakers. Fundamental frequency for S1 decreases slightly until the late 70's, then gradually increases. Fundamental

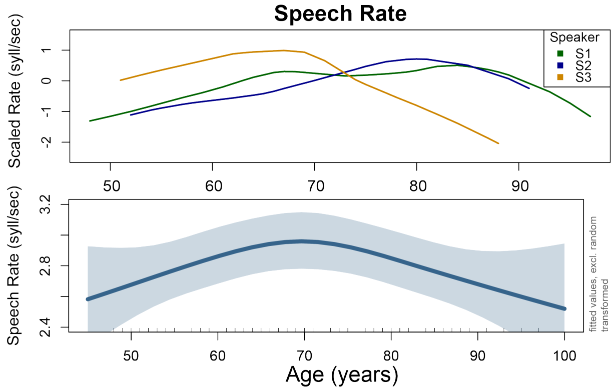

Figure 3: Changes in speech rate (syllables/second) for S1, S2, and S3 with advancing age (top) and Results of a GAMM analysis of speech rate over time (bottom)

frequency for S2 actually increases from 52 years to 59 years, then decreases until the mid 70's where, like S1, it begins to increase. S3 decreases very slightly until the early 70's and is followed by the sharpest increase out of all the speakers. As can be seen in Figure 4, modeled fundamental frequency follows a significant U-shaped pattern, decreasing from the age of around 50 years to 70 years, where it then follows an increasing trend with advancing age ($edf = 3.997, F = 298.528, p < 0.0001$).

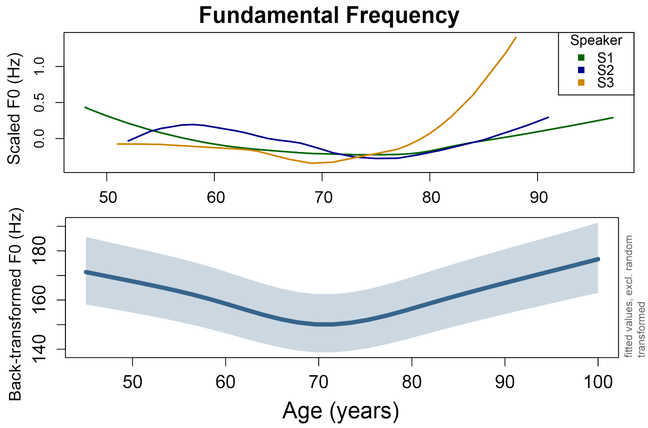

Figure 4: Changes in fundamental frequency for S1, S2, and S3 with age and GAMM of fundamental frequency for all speakers over time.

In Figure 5 (top left) we see S1's vowels space and how it changes over time. In this case, the vowel space shifts to lower F1 and F2 values. Changes in the vowel space of S2 can be seen in Figure 5 (top right). S2 also follows a similar trend as S1 although the shift is much less extreme. We note however that for S2 in his 50s there seem to be some surprisingly high F2 values. At the bottom of Figure 5, S3's vowel space is plotted, indicating that there was no shift to a decreased F1 but there may be a slight lowering of F2 over each decade. The vowel space area for speakers remains relatively constant with age (e.g., 1255 to 1366 for S1).

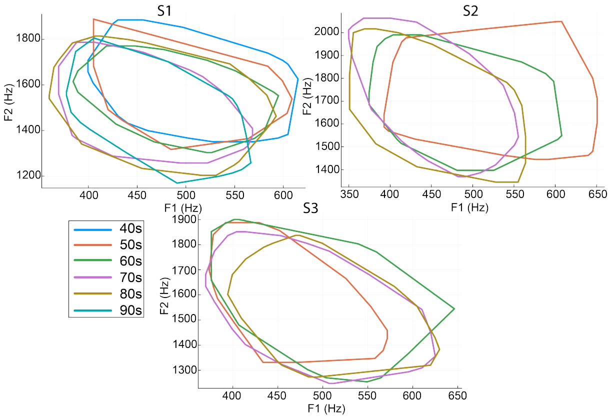

Figure 5: Overlaid vowel plots reflecting production for all three speakers over time. Each decade of the speaker's life is represented by a convex hull [27].

4. DISCUSSION AND CONCLUSION

This paper investigates the effects of age in three male speakers on segment and word duration, speech rate, fundamental frequency, and vowel space. Our data is longitudinal, controlling for potential confounding variables that arise with cross-sectional studies, such as individual differences. Our findings show that the acoustic changes that occur with age are dynamic and idiosyncratic. They are dynamic in that the acoustic changes do not follow a linear trend, but are more U-shaped (or upside-down U-shaped) with age. The results also indicate that each speaker presents individual patterns in their acoustic data.

Our results of vowel and word duration build on the previous literature. These results align with the literature in that segment and word duration is longer in late adulthood than in young adulthood. However, it adds to previous literature by suggesting that the trajectory follows a U-shaped pattern and not a strictly linear pattern.

There are interesting similarities between the trajectories of vowel duration and word duration. Both trajectories follow a U-shaped pattern where the segments and words are produced faster until around the age of 70 years when both start to be produced slower with advancing age. The relation of the two patterns makes sense as vowel duration ultimately affects whole word duration.

It is also important to note the similarities found between the two durational GAMMs and the speech rate GAMM. The two measures are related as segment and word duration influence speech rate. As vowel and word duration shorten, speech rate increases, and vice versa, as vowel and word duration lengthen, speech rate decreases as less syllables are able to be produced per second. In this study, the speech rate refers to specifically read speech given in a talk. Different results are expected in the cases of spontaneous speech or speech produced from a different elicitation tasks.

Speech rate follows an inverted U-shaped trajectory expected from results found by [6] who also found an inverted U-shaped pattern. However, [6] found the peak to occur at the age of 45 years, whereas our data peaks at around 71 years. This is likely due to the fact that our data only samples later years and doesn't include anything below 45 years old. Also, our results are different for speech rate than that found by [7] who found that speech rate increases in late adulthood. However, this strengthens Quene and colleagues' arguments that sociolinguistic factors contribute to their result.

Our findings that show a U-shaped trajectory for fundamental frequency support the findings of other researchers for males and females [14] and for Alistair Cooke [11]. However, a key difference between our results and the previous analyses is the age at which the U-shaped pattern occurs, as with speech rate. Our results suggest that on average for our three speakers there is a change in their early 70s. However, Reubold et al. [11] found that an increase in fundamental frequency occurred at the age of 87 years for Alistair Cooke, and Stathopoulos et al. [14] found it to be around 50 and 60 years for males and females respectively. This suggests that the age at which the change in the direction of fundamental frequency occurs is likely an indicator of individual aging processes and will be unique for every individual.

For vowel space, we find that in general for all of our male speakers, there is a decrease in F1 and F2 with aging. This corresponds with earlier findings [10, 11, 1, 16] but does not agree with findings by Gahl & Baayen who find a more peripheral vowel space with age. Again this may be due to differences is the sampled speakers' ages.

While the effects of age on vowel and word duration, speech rate, fundamental frequency, and vowel space from our data follow general trends that have been found in previous research, our results also indicate that there are dynamic and idiosyncratic patterns for individual speakers. Furthermore, the pattern of change is not linear. As a result, a statistical model that allows for non-linear trends is necessary for research on the acoustic changes related to aging.

5. REFERENCES

[1] S. A. Xue and G. J. Hao, "Changes in the human vocal tract due to aging and the acoustic correlates of speech production," *Journal of Speech, Language, and Hearing Research*, vol. 46, no. 3, pp. 689–701, 2003.

[2] B. V. Tucker, C. Ford, and S. Hedges, "Speech aging: Production and perception," *WIREs Cognitive Science*, vol. 12, no. 5, pp. 1–23, 2021.

[3] J. D. Harnsberger, R. Shrivastav, W. S. Brown, H. Rothman, and H. Hollien, "Speaking rate and fundamental frequency as speech cues to perceived age," *Journal of Voice*, vol. 22, no. 1, pp. 58–69, Jan. 2008.

[4] E. B. Slawinski, "Acoustic correlates of [b] and [w] produced by normal young to elderly adults," *The Journal of the Acoustical Society of America*, vol. 95, no. 4, pp. 2221–2230, Apr. 1994.

[5] M. Bruckl, "Women's vocal aging: A longitudinal approach," in *INTERSPEECH 2007*, Antwerp, Belgium, Aug. 2007, pp. 1170–1173.

[6] E. Jacewicz, R. A. Fox, and L. Wei, "Between-speaker and within-speaker variation in speech tempo of American English," *The Journal of the Acoustical Society of America*, vol. 128, no. 2, pp. 839–850, Aug. 2010.

[7] H. Quené, "Longitudinal trends in speech tempo: The case of Queen Beatrix," *The Journal of the Acoustical Society of America*, vol. 133, no. 6, pp. EL452–EL457, May 2013.

[8] S. N. Awan, "The aging female voice: Acoustic and respiratory data," *Clinical Linguistics & Phonetics*, vol. 20, no. 2-3, pp. 171–180, Jan. 2006.

[9] E. de Leeuw, "Native speech plasticity in the German-English late bilingual Stefanie Graf: A longitudinal study over four decades," *Journal of Phonetics*, vol. 73, pp. 24–39, Mar. 2019.

[10] J. Harrington, S. Palethorpe, and C. Watson, "Age-related changes in fundamental frequency and formants: A longitudinal study of four speakers," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 2, Jan. 2007, p. 2756.

[11] U. Reubold, J. Harrington, and F. Kleber, "Vocal aging effects on F0 and the first formant: A longitudinal analysis in adult speakers," *Speech Communication*, vol. 52, no. 7-8, pp. 638–651, Jul. 2010.

[12] A. Russell, L. Penny, and C. Pemberton, "Speaking fundamental frequency changes over time in women: A longitudinal study," *Journal of Speech, Language, and Hearing Research*, vol. 38, no. 1, pp. 101–109, Feb. 1995.

[13] J. T. Eichhorn, R. D. Kent, D. Austin, and H. K. Vorperian, "Effects of aging on vocal fundamental frequency and vowel formants in men and women," *Journal of Voice*, vol. 32, no. 5, pp. 644.e1–644.e9, Sep. 2018.

[14] E. T. Stathopoulos, J. E. Huber, and J. E. Sussman, "Changes in acoustic characteristics of the voice across the life span: Measures from individuals 4–93 years of age," *Journal of Speech, Language, and Hearing Research*, vol. 54, no. 4, pp. 1011–1021, 2011.

[15] E. J. Hunter, M. Kapsner-Smith, P. Pead, M. Z. Engar, and W. R. Brown, "Age and speech production: A 50-year longitudinal study," *Journal of the American Geriatrics Society*, vol. 60, no. 6, pp. 1175–1177, 2012.

[16] E. Jacewicz, R. A. Fox, and J. Salmons, "Cross-generational vowel change in American English," *Language Variation and Change*, vol. 23, no. 1, pp. 45–86, 2011.

[17] S. Gahl and R. H. Baayen, "Twenty-eight years of vowels: Tracking phonetic variation through young to middle age adulthood," *Journal of Phonetics*, vol. 74, pp. 42–54, May 2019.

[18] L. C. Cantor-Cutiva, P. Bottalico, and E. Hunter, "Factors associated with vocal fry among college students," *Logopedics Phoniatrics Vocology*, vol. 43, no. 2, pp. 73–79, Apr. 2018.

[19] L. A. R. Rossi-Barbosa, M. R. Barbosa, R. M. Morais, K. F. de Sousa, M. F. Silveira, A. C. C. Gama, and A. P. Caldeira, "Self-Reported Acute and Chronic Voice Disorders in Teachers," *Journal of Voice*, vol. 30, no. 6, pp. 755.e25–755.e33, Nov. 2016.

[20] H. Brugman and A. Russel, "Annotating multimedia / multi-modal resources with ELAN," *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, Mar. 2009.

[21] J. Yuan and M. Liberman, "Speaker identification on the SCOTUS corpus," *Proceedings of Acoustics '08*, pp. 5687–5690, 2008.

[22] P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer," 2019. [Online]. Available: www.praat.org

[23] T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models*. CRC Press, Jun. 1990.

[24] S. N. Wood, *Generalized Additive Models*. New York: Chapman & Hall/CRC, 2006.

[25] M. C. Kelley and D. Aalto, "Measuring the dispersion of density in head and neck cancer patients' vowel spaces: The vowel dispersion index," *Canadian Acoustics*, vol. 47, no. 3, pp. 114–115, Oct. 2019, number: 3.

[26] B. H. Story and K. Bunton, "Vowel space density as an indicator of speech performance," *The Journal of the Acoustical Society of America*, vol. 141, no. 5, pp. EL458–EL464, 2017.

[27] M. C. Kelley, "Phonetics.jl," Nov. 2021, original-date: 2020-05-27T04:51:46Z. [Online]. Available: https://github.com/maetshju/Phonetics.jl