Effects of Vocal Aging on Fundamental Frequency and Formants

S. Mwangi[1], W. Spiegl[2], F. Hönig[2], T. Haderlein[2], A. Maier[2], E. Nöth[2]

[1]*Electrical and Electronics Department, Jomo Kenyatta University of Agriculture and Technology*
P. O. Box 62000-00200 Nairobi, Kenya
[2]*Lehrstuhl für Mustererkennung, Universität Erlangen-Nürnberg*
Martensstraße 3, 91058 Erlangen, Germany

Abstract

In this paper we study changes of the articulatory organs from young adulthood to old age. In particular, we focus on how the larynx and vocal folds age along with the rest of the body. The objective of the paper is to automatically compute the effects of vocal aging on fundamental frequency and the formants.

We compare trends of the fundamental, 1st, 2nd, and 3rd formant frequencies as the age progresses. We also investigate how different phonemes vowels in particular are affected by vocal aging. The data consists of Queen Elizabeth II's annual Christmas speech from the age of 26 to 76. The idea is to have the same speaker over a long period of time.

The results indicate that vocal aging causes the decline of the fundamental and 1st formant frequencies while the 2nd and 3rd formants are not closely correlated to vocal aging. In general, the correlation between the fundamental frequency and age is -0.90, while the average correlation for the 1st formant and age is -0.81. In particular, the correlation between the 1st formant frequencies of phoneme 'o' and age is -0.89 while for the 2nd formant it is 0.34.

Index Terms: Prosody, Biomedical acoustics, Speech processing

1. Introduction

The larynx and vocal folds age along with the rest of the body. They also undergo age-related anatomic changes during adulthood. This leads to a change of the voice quality in several aspects, such as fundamental frequency and formants as shown by Harrington et al. [1, 2]. These changes typically result in a voice which is perceived as "old". Other respiratory system changes include stiffening of the thorax and weakening of respiratory muscles.

Knowledge on the effects of vocal aging on fundamental frequency and formants is important as it can be used to model aging in the design of speech recognition and speech synthesis systems. Speech engines can be designed to either predict the age of a speaker (age recognition; cf. [3]) or generate speech of a person of a certain age (synthesis) depending on the system's requirements. Especially for speech recognition, the predicted age can be used to automatically adjust the system parameters for suitable recognition parameters.

2. Approach

2.1. Speech Production

The vocal tract begins at the larynx (vocal cords or glottis) and ends at the lips. The nasal tract, which begins at the velum and ends at the nostrils, is relevant during the articulation of nasal sounds.

Speech is then produced in the following manner:

- Air is pushed from the lungs through the vocal tract, out of the mouth.
- For voiced sounds, the vocal cords vibrate (open and close) and create therewith a primary voice signal.
- For unvoiced sounds, the vocal cords do not vibrate but remain constantly opened.
- Depending on the shape of the vocal tract and the positions of the various articulators (i.e., jaw, tongue, velum, lips, mouth), different sounds are produced.

The source-filter model [4] is a simplified mathematical model of the speech production outlined above. The speech signal is the output of a filter whose input is either a sequence of impulses or a white noise. The parts of the source-filter model can be mapped to the parts of the vocal tract:

- **Vocal Tract** ⟶ *Linear filter*
- **Vocal Cord Vibration** ⟶ *Voiced sounds*
- **Vocal Cord Vibration Period** ⟶ *Pitch period*
- **No Vocal Cord Vibration** ⟶ *Unvoiced sounds*
- **Air Volume Flow** ⟶ *Gain*

Thus, the source-filter model describes speech as a combination of a sound source, such as the vocal cords, and a filter (cf. Figure 1).

2.2. Data

The data used for this project consisted of Queen Elizabeth II's annual Christmas speech over a period of 50 years. The data is courtesy of the British Broadcasting Cooperation (BBC). This data represents the speech of the same person over a long period of time. Due to that, the recording equipment was changed at several points. This has to be kept in mind in the subsequent steps of processing.

2.3. Evaluation Techniques

In order to isolate vowels in the speech, the signal was first segmented into voiced and unvoiced parts. The voiced and unvoiced speech segments are discriminated by using thresholding parameters [5]: short-time energy, zero-crossing rate, and maximum amplitude. The following methods were then used for fundamental frequency and formant detection:

- Autocorrelation method
- Linear Prediction Coding (LPC) [6, 7]

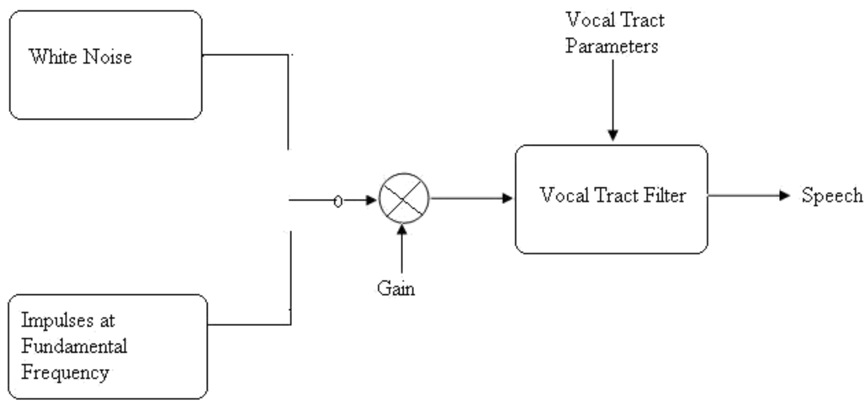

Figure 1: Diagram of the source-filter model (after [4])

For fundamental frequency detection, the autocorrelation method was chosen because it is less sensitive to noise in the speech signal than for example LPC. The speech acquisition equipment in the younger age of the Queen were vulnerable to noise, thus it was necessary to use this method. Short-time autocorrelation $R_n(k)$ of a windowed speech segment revealed the periodicity in voiced speech segments. This property was used to extract the fundamental frequency of the Queen's speech.

$$R_n(k) = \sum_{m=0}^{m=N-1-k} [x(n+m)w(m)][x(n+m+k)w(k+m)] \tag{1}$$

N is the window length, $x(n)$ the current sample, and $w(m)$ is the window function. For formant extraction, the LPC method was chosen for the source-filter model. In this method, we model the speech signal coming from a source through a filter and we extract the resonant frequencies using the *Linear Prediction* analysis [6, 7]. From the spectrum of the vocal tract transfer function, we derive the formant estimates. We used the implementation of the formant tracker as provided by WaveSurfer [8]. Using forced alignment, we segmented the phonemes.

3. Results and Discussion

3.1. Fundamental Frequency

As the body ages, the anatomy of the larynx changes, and therefore the vocal cords do not vibrate at the same frequency throughout. This was observed from the plot of fundamental frequency against age (cf. Figure 2). The fundamental frequency was averaged over each of the Queen's speeches. The general trend was a considerable decrease between the ages of 26 to 43 followed by a marginal decline between the ages of 43 to 70. We relate the changes in the fundamental frequency to changes of the tissue in the vocal cords. Whether this is dependent on hormonal changes during these periods is unclear. Then there is again a stronger decrease in the fundamental frequency from the age of 70 and above. The correlation between the fundamental frequency and age was -0.90, and hence the explained variance was 81 %. Thus, most of the decline in fundamental frequency was explained by age.

vowel	$r(F_1)$	$r(F_2)$	$r(F_3)$
/{/	-0.75	-0.05	0.39
/e/	-0.81	0.01	0.43
/I/	-0.86	0.08	0.40
/Q/	-0.89	0.34	0.61
/U/	-0.55	0.02	0.17
/A:/	-0.41	0.44	0.56
/i:/	-0.87	0.19	0.26
/O:/	-0.82	-0.04	0.23
/u:/	-0.79	0.71	0.30

Table 1: Overview on the correlations r between formant frequencies (first F_1, second F_2, and third formant F_3) and age for the different vowels in SAMPA notation [9].

3.2. Formants

Analysis of formants was carried out on different vowels. Note that the SAMPA annotation [9] is used in the following. An overview of the results is presented in Table 1. Figure 3 displays the first, second and third formants of vowel /e/ plotted against age. The correlation of the first formants and age was between -0.41 to -0.89 and the explained variance between 17 and 79 %. This meant that the decline of the first formant was explained by the age in most vowels. Only for vowels /U/ and /A:/, the first formant did not correlate strongly to age.

The correlation between the second formant and age was between 0.01 to 0.44 and the explained variance was between 1 and 19 %. Thus, only little of the increase in the second formants was explained by age. In the thirds formants, we also get variable correlations depending on the phoneme. The lowest correlation was 0.17 in /U/ and the highest 0.61 in /Q/. This corresponds to an explained variance between 3 and 37 %.

We can conclude that most of the first formants decrease with growing age of a female person as also observed by Harrington et al. [1, 2]. Usually, the first formant is explained by the height of the tongue. Hence the decrease of the first formants with growing age may be related to a relaxation of the tongue and its muscles with age.

The effect of the age on the second and third formant is vari-

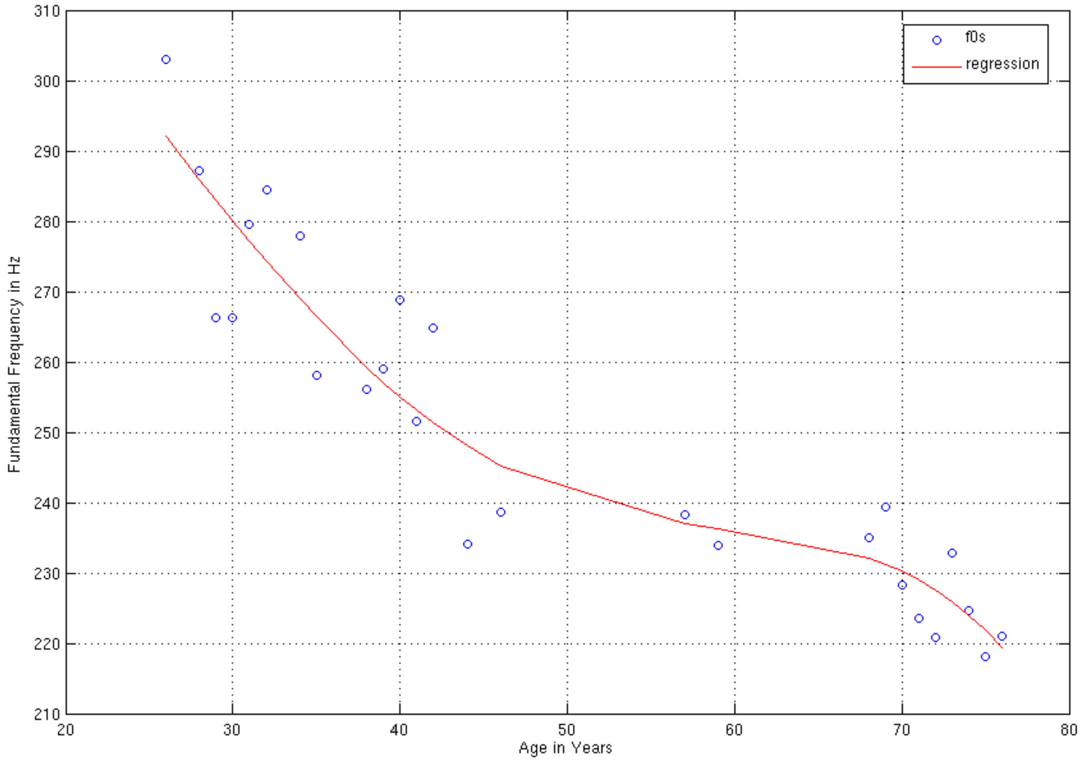

Figure 2: Fundamental frequency computed over all voiced segments against age.

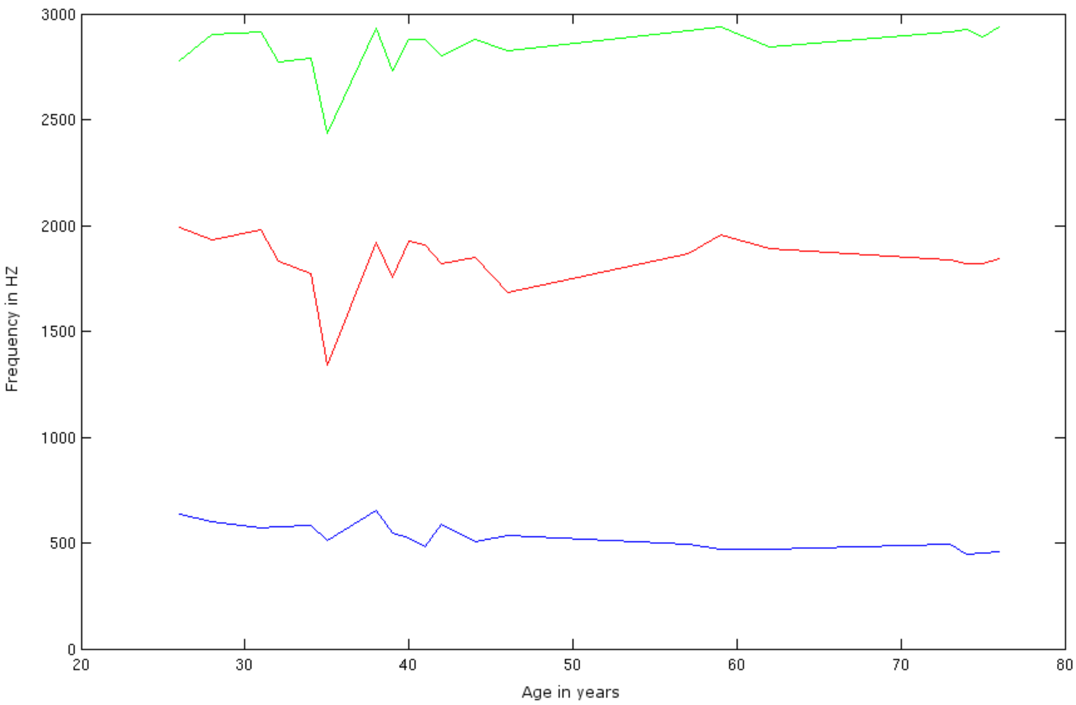

Figure 3: Formants of phoneme /e/ against age ($r(F_1) = -0.81$, $r(F_2) = 0.01$, $r(F_3) = 0.43$).

able. Only in few vowels, a high correlation was found. Mostly, there are only weak ties between age and the higher formant frequencies. These formants remain constant even with growing age. As the second formant is often related to the tongue position, it seems that the tongue position during articulation also remains stable with growing age.

4. Summary

In this paper we presented a longitudinal study on vocal aging. We investigated speech recordings of a single female speaker recorded over 50 years annually. It could be shown that there is a strong negative correlation between the fundamental frequency of the speaker and her age.

Furthermore, we investigated the formants of different vowels during the different ages. For most phonemes, the first formants decreased with growing age. The second and third formants remained rather unaffected.

5. Acknowledgment

This research project is supported by the BFS (Bayerische Forschungsstiftung) research association "Zukunftsorientierte Produkte und Dienstleistungen für die demographischen Herausforderungen — FitForAge" (www.fit4age.org). We thank Jonathan Harrington and the BBC for providing the recordings of the annual Royal Christmas Message of Queen Elisabeth II.

Furthermore, we would like to thank all participants of the Johns Hopkins University (JHU) Summer Workshop 2008 "Vocal Aging Explained by Vocal Tract Modeling" who have contributed to this work.

6. References

[1] J. Harrington, "An acoustic analysis of 'happy-tensing' in the Queens's Christmas broadcasts," *Journal of Phonetics*, vol. 34, pp. 439–457, 2006.

[2] J. Harrington, S. Palethorpe, and C. I. Watson, "Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers," *Interspeech 2007*, pp. 2753–2756, 2007.

[3] T. Bocklet, A. Maier, J. Bauer, F. Burkhardt, and E. Nöth, "Age and Gender Recognition for Telephone Applications based on GMM Supervectors and Support Vector Machines," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1. Las Vegas, USA: IEEE Computer Society Press, 2008, pp. 1605–1608.

[4] G. Fant, *Acoustic Theory of Speech Production*. The Hague, The Netherlands: Mouton, 1960.

[5] A. Kießling, *Extraktion und Klassifikation prosodischer Merkmale in der automatischen Sprachverarbeitung*, ser. Berichte aus der Informatik. Aachen, Germany: Shaker, 1997.

[6] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, April 1975.

[7] J. Markel and J. A.H. Gray, *Linear Prediction of Speech*. Springer Verlag, 1976.

[8] K. Sjolander and J. Beskow, "WaveSurfer Homepage," 2006, http://www.speech.kth.se/wavesurfer/, last visited 03/08/2009.

[9] J. Wells, *SAMPA computer readable phonetic alphabet*. Berlin, Germany: Mouton de Gruyter, 1997, ch. Part IV, section B.