

# Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers.

Jonathan Harrington<sup>1</sup>, Sallyanne Palethorpe<sup>2</sup>, Catherine I. Watson<sup>3</sup>

<sup>1</sup> Institute of Phonetics and Speech Processing (IPS), University of Munich, Munich, Germany.

<sup>2</sup> Macquarie Centre for Cognitive Science (MACCS), Macquarie University, Sydney, Australia.

<sup>3</sup> Dept. of Electrical and Computing Engineering, The University of Auckland, New Zealand.

jmh@phonetik.uni-muenchen.de, sassi@maccs.mq.edu.au, c.watson@auckland.ac.nz

## Abstract

The study is concerned with a longitudinal acoustic analysis of two sets of recordings from the same four speakers over an interval of between 29 and 50 years. The aim was to determine whether there is any evidence for age-related acoustic changes. Our analysis showed that the same speakers have lower  $f_0$ , a lower F1, a marginally lower F2, and an unchanging or sometimes higher F3 in their later recordings. There is some suggestion from these data that the change in F1- $f_0$  in Bark from earlier to late recordings is proportional to the change in F3-F2 in Bark. This suggests that there is shift in the speaker space roughly along a diagonal in the phonetic height x backness plane with increasing age.

## 1. Introduction

In recent studies of the Christmas broadcasts of Queen Elizabeth II [1-4], we have found evidence of two kinds of changes. The first is for a rearrangement of the relative position of some of the vowels that is compatible with findings from other acoustic and auditory analyses of vowel change (e.g., [5,6]) in the standard accent of England, Received Pronunciation (RP). Thus in both the data from the Queen and in these RP studies there is evidence that the vowel exemplified by the lexical set in TRAP has become more open and that the vowel in GOOSE has fronted in the last 50 years. Moreover, the evidence that these vowel changes were in the direction of 1980s RP speakers – i.e., that the Queen's 1980s vowels occupy a position intermediate between her 1950s vowels and those of five female RP speakers from the 1980s [2,3] – seems to be further evidence that we are dealing with subtle *phonetic* changes within the lifespan of the same adult.

The studies of the Christmas broadcasts in [1] also provided evidence for age-related changes that may come about because of physiological maturation of the vocal tract with increasing age. More specifically, in an investigation of schwas in the Christmas broadcast data [1], the fundamental frequency decreased from an average of 253 Hz to 189 Hz and F1 decreased from 551 Hz to 428 Hz. There were also changes in the higher formants: a drop in F2 from 1696 Hz to 1624 Hz, but an increase in F3 from 2834 Hz to 2864 Hz. Similar changes in these parameters were found for schwa within the same word (in the second syllable of 'Christmas') thus ruling out the possibility that these differences in the acoustics of schwa across these decades could be an artifact of different contexts. Since there is also no reason to expect that we are dealing with a phonetic change to schwa and since neither these schwa changes nor the changes to the vowels of the Christmas broadcasts in general could be explained as a result of speaking style changes [1,2,4], the more likely

interpretation is that the shift in the schwa formants is a correlate of vocal tract changes due to age.

Other studies in which the relationship between age and vowel acoustics has been examined by comparing different groups of young and old speakers are compatible with some of these findings. The drop in  $f_0$  with increasing age is fairly well documented [7-11] and there is some more evidence that F1 is lower in older than younger speakers [9].

In view of the paucity of acoustic studies dealing with the relationship between age and formants, we undertook a further longitudinal analysis in three other speakers for which recorded materials from broadcasts were available spanning at least 35 years. The speakers included an RP speaker, Margaret Lockwood, a British actress; and an Australian and New Zealand English speaker. We chose Margaret Lockwood because she is one of the very few female RP speakers for which broadcast data is available over a long period of time (thereby allowing comparisons to be made with the Queen). The Australian and New Zealand English male speakers were chosen partly because sound change in Australian and New Zealand English is being investigated in other projects by the second and third authors respectively of this paper; but also because if these schwa-changes are really non-phonetic and age-dependent, then they should carry over to male speakers and to other dialects.

## 2. Method

### 2.1. Speakers and materials

The speech data from Queen Elizabeth II (born 1926) were taken from the annual Christmas broadcasts in the following decades: 1950s (52, 54-59); 1960s (60, 62, 63-68); 1970s (70-72); 1980s (83, 85, 88); 1990s (94-99); 2000s (00-02). The average duration of the broadcast data was 5.2 minutes; the shortest broadcast was 1.0 minutes (1959) and the longest was 8.2 min (1956). The recordings from Margaret Lockwood (born 1916) were from 1951 (duration 5.5 minutes) and 1980 (duration 12.0 minutes) and were both taken from the same, long-running BBC radio program 'Desert Island Discs'. The recordings from the Australian English (AE) male speaker (born 1908) are from 1948 (duration 23.0 minutes) and 1987 (duration 29.0 minutes). The 1948 data were from a speech at the Cricket Writers' Club Dinner; the 1987 data were from an interview with the Australian Broadcasting Corporation. The speech recordings from the New Zealand (NZ) English speaker (born 1919) are from 1955 (duration 14.1 minutes) and 1992 (duration 14.2 minutes). Both recordings are from the Radio New Zealand Sound Archives. The 1955 recording is a monologue broadcast aimed at school children, outlining an upcoming expedition to the Antarctic. The 1992 recording

is a radio interview in which the speaker recounts details of his life to an interviewer.

We labelled the data from these additional three speakers following the procedures in [1] in which 544 and 807 of the Queen's schwas in broadcasts from 1952-60 and from 1995-2002 respectively had been labelled. As in [1], schwas were labelled in polysyllabic content words and care was taken not to include those weak vowels that are realized with a weak [ɪ] in RP that is, schwas were only taken from weak vowels transcribed in most dictionaries with /ə/ as in 'Rosa's' (but not 'roses'). Schwas were not included if their duration was so short that no reliable formant frequencies could be calculated. The number of schwas labeled were as follows: for Lockwood 82 from 1951, and 183 from 1980; for the AE speaker, 333 from 1948, and 371 from 1987; for the NZ speaker, 219 from 1955, and 247 from 1992.

The speech data from the Christmas broadcasts were digitised at 16 kHz except those from 1952, 54, 57, 83, 85, 88 which had been digitised at 20 kHz. The data from Lockwood were digitised at 16 kHz and those from the AE and NZ speakers at 20 kHz. The first four formants were calculated using the default method in the tkasp-EMU signal processing software using an LPC-based autocorrelation method and the split Levinson algorithm. A Blackman window was used and the window shift and size were set to the default of 5 ms and 30 ms.  $f_0$  values were calculated with the same 5 ms frame shift.

We extracted the first three formant frequencies and the fundamental frequency for all these data at the schwa midpoint. In addition, we also extracted  $f_0$  and F1-F3 for all frames of voiced speech in the entire recordings and averaged  $f_0$  and formants separately by year for each speaker; we did this because we felt that a comparison across all voiced frames would also give some indication of age-related  $f_0$  and F1-F3 shifts in the speaker's space.

### 3. Results

#### 3.1. Christmas broadcast data

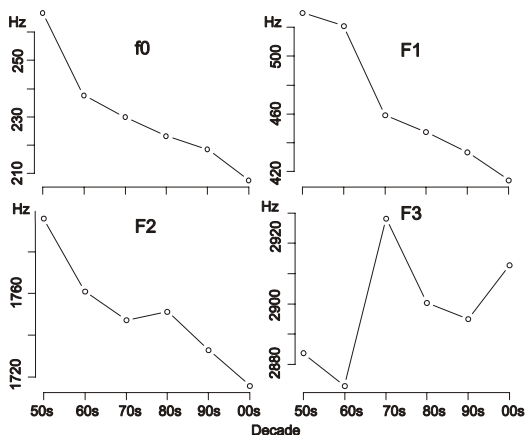


Fig. 1.  $f_0$  and F1-F3 values averaged across voiced frames separately by decade in the Queen's Christmas broadcasts.

Fig. 1 shows the average values of  $f_0$  and F1-F3 separately by decade between the 1950s and 2000s across all the frames of voiced speech data. There is very clearly a trend for both  $f_0$  and F1 to fall from earlier to later decades. The extent of the fall in  $f_0$  is 59 Hz (22.1%, calculated with (1) below) from 267 Hz to 208 Hz and that of F1 is 116 Hz (21.9%) from 530 Hz to 414 Hz. There is also a tendency for F2 to fall: the

extent of the fall from the 1950s to the 2000s is 80 Hz (4.5%) from 1796 Hz to 1716 Hz. Although there is an increase in F3 from the 1950s to the 2000s, Fig. 1 shows that there is much less evidence of a trend across the decades. Moreover, the extent of the rise between the 1950s and 2000s in F3 is small at 29 Hz (1.0 %).

We carried out a two-sample  $t$ -test in which we compared the average values from the 1950s (7 average values, one each for 1952, 1954-59) with the average values from the 1990s and 2000s together (9 average values, 1994-2002). The results showed a significant difference on  $f_0$  ( $t = 7.2$ ,  $df = 9.4$ ,  $p < 0.001$ ) on F1 ( $t = 11.2$ ,  $df = 10.7$ ,  $p < 0.001$ ) and on F2 ( $t = 2.83$ ,  $df = 8.6$ ,  $p < 0.05$ ), but not on F3.

#### 3.2. Margaret Lockwood

Fig. 2 (left) shows the average proportional change in Hz across all frames of voiced data from the early (1951) to the later (1980) broadcasts. The averages were calculated from (1):

$$100 (f_L - f_E)/f_E \quad (1)$$

where  $f_L$  and  $f_E$  are the averages on that parameter ( $f_0$ , F1-F3) in the late and early broadcasts respectively.

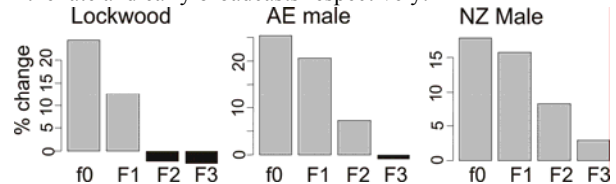


Fig. 2. The mean percentage change in  $f_0$  and F1-F3 obtained from averages across all voiced frames for three speakers. Positive values on (1) are shown by grey filled bars in which the averages were higher in the earlier broadcast i.e., there is a decrease on that parameter from the early to the late recording. Negative values on (1) are shown by black filled bars.

Fig. 2 (left) shows a decrease in  $f_0$ , F1 and a very modest increase in F2 and in F3 from early to late broadcasts. The data extracted at the temporal midpoint of the schwa are generally in the same direction, but the extent of the change is much less: at the schwa midpoint,  $f_0$  falls from 156 Hz to 123 Hz; F1 falls from 481 Hz to 474 Hz; F2 falls from 1518 Hz to 1503 Hz; and F3 rises from 2468 Hz to 2557 Hz from earlier to later broadcasts. Only the change in  $f_0$  is significant, however ( $t = 6.5$ ,  $df = 121.7$ ,  $p < 0.001$ ).

#### 3.3. Australian English male speaker

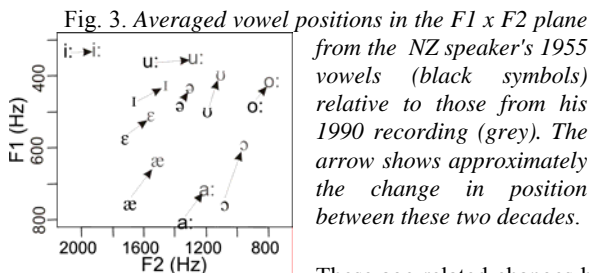
Across all frames of voiced data, the pattern of change from the AE speaker's early (1948) to the later (1987) broadcast is similar to those for the Queen: as Fig. 2 (centre) shows, there is a decrease in  $f_0$ , F1, and F2 while F3 increases from the early to the late broadcast. The data from the temporal midpoint of the AE speaker's schwas show the same trend. The mean values in the early broadcast are: 166 Hz ( $f_0$ ), 535 Hz (F1), 1578 Hz (F2), 2510 Hz (F3); the corresponding mean values in the late broadcast are: 123 Hz ( $f_0$ ), 420 Hz (F1), 1429 Hz (F2), 2594 (F3). The results of a  $t$ -test comparing schwas in the early with the late broadcast showed significant differences on all these parameters ( $f_0$ :  $t = 17.8$ ,  $df = 428.2$ ,  $p < 0.001$ ; F1:  $t = 15.6$ ,  $df = 595.5$ ,  $p < 0.001$ ; F2:  $t = 10.7$ ,  $df = 558.9$ ,  $p < 0.001$ ; F3  $t = -6.2$ ,  $df = 450.1$ ,  $p < 0.001$ ).

### 3.4. New Zealand English speaker

The percentage change in the mean fundamental and formant values from the early (1955) to the late (1990) broadcast across all frames of data is shown in Fig. 2 (right). Fig. 2 shows a decrease on all parameters from the early to the late broadcasts. The same trend is evident in the averaged parameters extracted at the temporal midpoint of the schwa: in the early broadcast, the mean values on  $f_0$ , and F1-F3 respectively are 131 Hz, 479 Hz, 1320 Hz, and 2265 Hz; the corresponding averages for the late broadcast are: 111 Hz, 429 Hz, 1259 Hz, 2225 Hz. There is a significant difference between the early and the late broadcast on all these parameters: for  $f_0$ :  $t = 10.3$ ,  $df = 408.5$ ,  $p < 0.001$ ; for F1:  $t = 5.3$ ,  $df = 400.2$ ,  $p < 0.001$ ; for F2:  $t = 3.4$ ,  $df = 448.2$ ,  $p < 0.001$ ; for F3:  $t = 2.4$ ,  $df = 451.7$ ,  $p < 0.05$ .

## 4. Discussion

The findings from this study can be summarized as follows. Firstly, and compatibly with other studies, there is a decrease in  $f_0$  from the early to the late speech recordings in all four speakers. Secondly, all four speakers showed a decrease in F1. Thirdly, there was a tendency for F2 to fall, although the fall was not as marked as for  $f_0$  or F1. Fourthly, the F3 changes were least consistent, although in three speakers there was a tendency for F3 to rise from the earlier to the later recordings. Finally, the results were usually comparable across the two types of analysis, i.e. similar changes from early to late recordings were found in comparing averages across all voiced frames with schwas.



These age-related changes have some implications for diachronic phonetic changes to vowels. Consider the data in Fig. 3 which shows the averaged positions of various vowels in the formant plane from these two broadcasts by the NZ speaker. Since most of the changes are in the direction of those found in the present study for schwa and the voiced frames – i.e., a lowering of both F1 and F2 – it would be premature and probably incorrect to infer that most of the shifts in Fig. 3 were evidence of phonetic changes. The evidence for a phonetic shift in the NZ speaker's vowel data must therefore depend on a change in the relative position of the vowels: for example, some of F2-lowering of [ɪ] (HID) in Fig. 3 could well be evidence for the kind of phonetic-backing found that has been in other studies of New Zealand English [12-15], because, as argued in [15], the 1990s [ɪ] is proportionately nearer to [ə] than in the 1950s recording.

We now consider three kinds of explanation for  $f_0$  and formant changes shown in Fig. 2. The first is purely acoustic and trivial: the F1-lowering may just be an artifact of the LPC or formant tracking algorithm. For example, since the ratio in Hz between F1 and  $f_0$  in the Queen's broadcast data is more or less constant across the decades at roughly 2, then perhaps

the calculated F1 is simply tracking the 2<sup>nd</sup> harmonic: i.e., as  $f_0$  lowers across the decades, so does the 2<sup>nd</sup> harmonic and therefore so does F1. On the other hand, the data from the other speakers do not show such constancy. For example across all voiced frames, the F1/ $f_0$  Hz ratio for Lockwood increases from 2.91 to 3.38, while the corresponding increases for the AE and NZ speakers are from 3.01 to 3.30 and from 3.40 to 3.48 respectively.

The second interpretation is couched in articulatory terms: perhaps, as Laver & Trudgill [16] have suggested, the respiratory system and digestive tract are in a lower position resulting in vocal tract lengthening with increasing age. However, while this interpretation is compatible with F1 and F2 lowering, the data of the present study suggesting no change, or an increase, in F3 are not.

The third explanation has an auditory-perceptual basis: perhaps both  $f_0$  and F1 decrease in order that the perceived phonetic height is maintained with increasing age, given Traummüller's [17,18] experiments and model showing that the difference between F1 in Bark and  $f_0$  in Bark (henceforth  $F1_{\text{Bark}} - f0_{\text{Bark}}$ ) is a perceptual cue for vowel height. However a problem with this explanation is that  $F1_{\text{Bark}} - f0_{\text{Bark}}$  is generally around 0.5 Bark greater in the speakers' earlier recordings than in their later ones: that is, the F1 drop is greater than the  $f_0$  drop with the result that there is a convergence of  $F1_{\text{Bark}}$  and  $f0_{\text{Bark}}$ . But this would imply that the vocal tract is slightly more constricted (i.e., phonetically higher/closer) in later recordings.

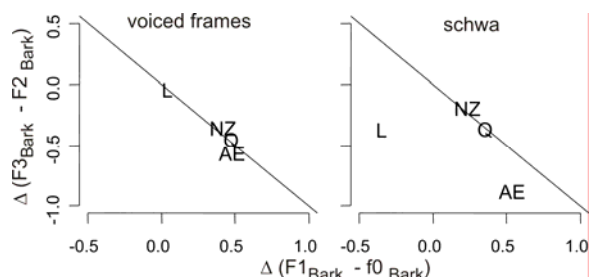


Fig. 4. The average change from early to late recordings in  $F3_{\text{Bark}} - F2_{\text{Bark}}$  as a function of the average change in  $F1_{\text{Bark}} - f0_{\text{Bark}}$  in voiced frames (left) and schwas (right) for the four speakers with the line  $y = -x$  superimposed on the data.

At the same time, we also have to account for the changes to F2 and F3. Interestingly, there is an approximate relationship between the change in  $F3_{\text{Bark}} - F2_{\text{Bark}}$  and the change in  $F1_{\text{Bark}} - f0_{\text{Bark}}$  just considered (the former following Syrdal & Gopal [19] is a cue for phonetic backness: front vowels have F3 and F2 closer together and hence a lower value on this parameter than back vowels).

We calculated for each of the four speakers  $\Delta(F3_{\text{Bark}} - F2_{\text{Bark}})$  the difference between the mean value of  $F3_{\text{Bark}} - F2_{\text{Bark}}$  in a speaker's early recording and the mean value of  $F3_{\text{Bark}} - F2_{\text{Bark}}$  in the same speaker's late recording. We obtained  $\Delta(F1_{\text{Bark}} - f0_{\text{Bark}})$  in an analogous way. A plot of these two parameters against each other (Fig. 4) shows that the change in  $F3_{\text{Bark}} - F2_{\text{Bark}}$  is in the opposite direction to the change in  $F1_{\text{Bark}} - f0_{\text{Bark}}$ , at least for three of the speakers. Essentially, Fig. 4 means that while F1 and  $f_0$  converge, F3 and F2 diverge by roughly the same Bark quantity from the early to the later recordings. But this would imply that the speaker space is phonetically closer and backer with increasing age.

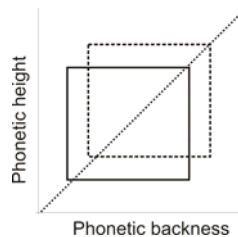


Fig. 5. *The proposed model of the shift in the speaker's space in the height x backness plane with increasing age (the dotted rectangle represents the same speaker's space at an older age).*

However, this is not at all the interpretation that we wish to offer and in order to explain why this is so, we need to abstract further away from these data and consider the speaker space as a region that occupies the phonetic height x backness plane, as in Fig. 5. Essentially, our data suggest that the spaces of the same speaker at young and older ages fall along a diagonal (in fact along the line  $y = x$  given that the change in  $F1_{\text{Bark}} - f0_{\text{Bark}}$  is offset by the change in  $F3_{\text{Bark}} - F2_{\text{Bark}}$ ): that is, with increasing age, the speaker's space moves up slightly along the diagonal towards the top right of the plane. Such a relationship is very reminiscent of the figures and data in Potter & Steinberg [20] and Peterson [21] who suggested that the relationship between different younger and older speakers might be connected by just such lines – and moreover that listeners might use these constant ratios to factor out, or normalize for, these speaker differences. So it seems that the age-related changes in the *same* speaker may mimic (but to a much smaller degree) the differences across *different* young and old speakers. This is why we would argue that this shift along the diagonal does not produce a change in phonetic quality: that is, if listeners learn to factor out such shifts across *different* young and old speakers in order to attain phonetic constancy, then so too should a shift of the *same* speaker's space along the diagonal remain phonetically neutral.

It remains an open and interesting question whether the shifts that we have seen in these data are the result of passive physiological changes to the vocal tract, or whether speech production is actively modified with increasing age, in order to compensate perceptually for the influences of the age-related decline in  $f0$  on vowel quality.

## 5. References

- [1] Harrington, J. "An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts", *Journal of Phonetics*, 34, 439–457, 2006.
- [2] Harrington, J., Palethorpe, S. and Watson, C. "Deepening or lessening the divide between diphthongs? An analysis of the Queen's annual Christmas broadcasts", in W.J. Hardcastle & J. Beck (eds.). *The Gift of Speech*. 2005. Erlbaum. (p. 227-261).
- [3] Harrington, J., Palethorpe, S., and Watson, C. "Does the Queen speak the Queen's English?" *Nature*, 408, 927-928, 2000.
- [4] Harrington, J. "Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts". In J. Cole & J. Hualde (eds.) *Laboratory Phonology 9*. Mouton: Berlin. In press.
- [5] Bauer, L. "Tracing phonetic change in the received pronunciation of British English". *Journal of Phonetics*, 13, 61-81, 1984.
- [6] Hawkins, S. & Midgley, J. "Formant frequencies of RP monophthongs in four age groups of speakers". *Journal of the International Phonetic Association*. 35, 183–199. 2005.
- [7] Decoster, W., & Debruyne, F. "Longitudinal voice changes: facts and interpretation". *Journal of Voice*, 14, 184–193, 2000.
- [8] Linville, S. *Vocal aging*. San Diego: Singular Thomson Learning. 2001.
- [9] Whiteside, S. "Sex-specific fundamental and formant frequency patterns in a cross-sectional study". *J. Acoust. Soc. Am.*, 110, 464–478, 2001.
- [10] Helfrich, H. "Age markers in speech". In K. Scherer & H. Giles (Eds.) *Social markers in speech*. (pp. 63-107). Cambridge: Cambridge University Press. 1979.
- [11] Linville, S. & Fisher, H. "Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females". *J. Acoust. Soc. Am.*, 78, 40-48, 1985.
- [12] Maclagan, M. & Hay, J. "Getting *fed* up with our *feet*: Contrast maintenance and the New Zealand English "short" front vowel shift". *Language Variation and Change*, 19, 1-25, 2007.
- [13] Watson, C., Maclagan, M. & Harrington, J. "Acoustic evidence for vowel change in New Zealand English". *Language Variation and Change* 12: 51–68, 2000.
- [14] Bauer, Laurie. "The second great vowel shift?" *Journal of the International Phonetic Association* 9: 57–66, 1979.
- [15] Watson, C., Palethorpe, S., Harrington, J.. "Capturing the vowel change in New Zealand English over a thirty year period via a diachronic study." *Proceedings of the 10th Australian International Conference on Speech Science & Technology* Sydney, p201-206, 2004.
- [16] Laver, J., & Trudgill, P. "Phonetic and linguistic markers in speech". In K. Scherer, & H. Giles (eds.), *Social markers in speech* (pp. 1–26). Cambridge: Cambridge University Press. 1979.
- [17] Traunmüller, H. "Perceptual dimension of openness in vowels". *J. Acoust. Soc. Am.*, 69, 1465–1475. 1981.
- [18] Traunmüller, H. "Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels". *Speech Communication*, 3, 49–61, 1984.
- [19] Syrdal, A. & Gopal, H. "A perceptual model of vowel recognition based on the auditory representation of American English vowels". *J. Acoust. Soc. Am.*, 79, 1086-1100, 1986.
- [20] Potter, R. K., and Steinberg, J. C. "Toward the specification of speech," *J. Acoust. Soc. Am.* 22, 807-820, 1950.
- [21] Peterson, G. E. "Parameters of vowel quality," *J. Speech Hear. Res.* 4, 10-29, 1961.