Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 52 (2010) 638–651
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

Vocal aging effects on F_0 and the first formant: A longitudinal analysis in adult speakers

Ulrich Reubold *, Jonathan Harrington, Felicitas Kleber

Institute of Phonetics and Speech Processing (IPS), University of Munich, Munich, Germany

Received 4 December 2009; received in revised form 4 February 2010; accepted 13 February 2010

Abstract

This paper presents a longitudinal analysis of the extent to which age affects F_0 and formant frequencies. Five speakers at two time intervals showed a clear effect for F_0 and F_1 but no systematic effects for F_2 or F_3. In two speakers for which recordings were available in successive years over a 50 year period, results showed with increasing age a decrease in both F_0 and F_1 for a female speaker and a V-shaped pattern, i.e. a decrease followed by an increase in *both* F_0 and F_1 for a male speaker. This analysis also provided strong evidence that F_1 approximately tracked F_0 across the years: i.e., the rate of change of (the logarithm of) F_0 and F_1 were generally the same. We then also tested that the changes in F_1 were not an acoustic artifact of changing F_0. Perception experiments with the main aim of assessing whether changes in F_1 contributed to age judgments beyond those from F_0 showed that the contribution of F_1 was inconsistent and negligible. The general conclusion is that age-related changes in F_1 may be compensatory to offset a physiologically induced decline in F_0 and thereby maintain a relatively constant auditory distance between F_0 and F_1.

Keywords: Vocal aging; F_0; Formants; Longitudinal analysis; Perception; Source–tract-interaction

1. Introduction

In the production of speech, several kinds of information are produced in parallel, both linguistic and also many that index personal attributes of the speaker. The latter certainly include cues that are linked to the physical properties of the speaker's vocal organs, including gender and the speaker's age range. In addition, many non-physiological speaker-specific characteristics leave their imprint on the acoustic signal, such as information about the speaker's emotional state as well as regional and social affiliation (see e.g. Chambers et al., 2002; Labov, 1972, 1994, 2001; Docherty, in press; Mendoza-Denton, in press, for recent reviews). The task of quantifying the relative contributions from these various sources of information is made especially difficult because many of them are indexed by similar sets of cues. For example, fundamental frequency depends both on the relationship between the pragmatic–semantic content of the utterance and intonation (Beckman and Venditti, 2010) and many other factors, such as the speaker's gender, age group, and emotional state (Campbell, 2004). Similarly, social and regional attributes tend to be parasitic upon the attributes of the speech signal that provide linguistic information (Labov, 2001): thus, the extent of glottalisation of a /t/ can provide information about syllable-affiliation in English but it can in some varieties also be a marker of social group affiliation (e.g., Foulkes and Docherty, 2006). Similarly, a more peripheral than usual high front lax vowel associated with a raising of the second formant frequency can be a cue that the word was produced with prosodic prominence (de Jong, 1995; Harrington et al., 2000a) but is also a marker of an Australian as opposed to a (Standard) British English variety (Cox, 2006).

* Corresponding author. Address: Institute of Phonetics and Speech Processing (IPS), University of Munich, Schellingstraße 3/II, 80799 Munich, Germany. Tel.: +49 8921802811; fax: +49 8921805790.
E-mail address: reubold@phonetik.uni-muenchen.de (U. Reubold).

0167-6393/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2010.02.012

In this paper, our main concern is with the effect of increasing age during the lifespan of an adult speaker on formant frequencies. Our interest in this association between formants and age is not only to advance our understanding of how increasing age in adulthood influences speech production, but also to assess the extent to which age-related formant changes may obscure phonetic differences between two groups of different aged speakers in the context of sociophonetic investigations. For example, sound change in progress can often be quantified in so-called apparent-time studies by comparing older with younger speakers of the same community (Bailey et al., 1991; Weinreich et al., 1968). However, the possible effect of non-phonetic age-related changes on formant frequencies is not usually taken into account in such studies (Harrington, 2006): consequently, apparent-time studies may under- or overestimate sound change if its effect on formants is different from, or similar to, that of physiologically-based changes due to ageing.

Some studies based almost entirely on the comparison of speakers in different age groups have shown an effect of age on formants and the most consistent effect is an age-dependent F_1-lowering (Linville and Fisher, 1985a; Scukanec et al., 1991; Xue and Hao, 2003). In addition, Rastatter et al. (1997) have shown a greater centralization of the vowel space in older speakers, although this effect was more marked in men than in women. The decreases of formants have been attributed to a lengthening of the vocal tract, caused by a lowering of the larynx, of the tracheobronchial tree and of the lungs, and by a growth of the facial skeleton (Laver and Trudgill, 1979; Linville, 2001). However, results are somewhat contradictory as far as the relative position of the larynx to the cervical vertebrae are concerned: while Zemlin (1998) cites Wind's (1970) findings of a vertical descent of the larynx during life, with a mean position of the lower border of the cricoid cartilage to cervical vertebrae descending from the 6th vertebra at age 20 to the 7th at age 80, Flügel and Rohen (1991) were not able to find a more lowered larynx during later adulthood in their study; and while Xue and Hao's (2003) analysis using an acoustic reflection technique found no differences between young (18–30 years) and old (62–79 years) speakers in overall vocal tract length, they did report both an increase in vocal tract volume, and in length and volume of the older speakers' mouth cavity.

In this paper, we are also concerned with an analysis of fundamental frequency, primarily because of the evidence that age-dependent F_0 and formant changes may be related (Harrington et al., 2007). Previous studies and compilations of studies have shown a decrease in F_0 with increasing age in women (Baken, 2005; Linville, 1996; Nishio and Niimi, 2008) and there is also some evidence for an age-dependent decrease in F_0 from longitudinal studies carried out in the same person (De Pinto and Hollien, 1982; Harrington et al., 2007; Russell et al., 1995; see also Mwangi et al., 2009). The comparable studies in male speakers are, however, less straightforward: F_0 is sometimes reported to remain unchanged (Verdonck-de Leeuw and Mahieu, 2004), to decrease (Decoster and Debruyne, 2000; Harrington et al., 2007) or to increase (Harnsberger et al., 2008). This different pattern of results may come about because of other evidence showing an F_0 drop in males over the age range 30–50 and then an increase in F_0 in older age (Baken, 2005; Linville, 1996). The F_0 decrease in women and the F_0 rise in later life in males have been attributed to the influence of hormonal changes on vocal fold vibration (Linville, 1996; see also Abitbol et al., 1999; Gugatschka et al., in press), while the F_0-decrease in both genders has also been explained in terms of an increase in vocal fold thickness and mass (Hollien and Shipp, 1972).

One of the reasons why studies of age-dependent influences of acoustic parameters may have provided conflicting results is because they are usually studied across different speaker groups (e.g., one young, one old): thus the different vocal tract morphologies which have a marked influence on F_0 and formants may obscure changes due to age. Similarly, the sociophonetic differences between younger and older speakers of the same speaking community may mask some of the physiologically based differences between them. One of the aims of the present study is to factor out some of these phonetic variables by carrying out a longitudinal study of age effects within the same speakers over a number of years. These types of studies are in general quite rare because of the difficulty of sampling materials over several decades from the same speaker. Moreover, even if such material is available, it is very often confounded by speaking style differences.

Although some recent longitudinal studies have been concerned with phonetic changes in the context of sound change in progress (Harrington et al., 2000b,c, 2005; Harrington, 2007; Sankoff and Blondeau, 2007), there is a paucity of those examining the non-phonetic long-term effects of age. As far as fundamental frequency is concerned, existing longitudinal studies are consistent with those based on between group comparisons in showing an F_0 decrease with increasing age in the same speaker (Decoster and Debruyne, 2000; Harrington, 2006; Russell et al., 1995). A longitudinal study by Endres et al. (1971) indicated age-dependent formant decreases, although since their results were based on an average across all formants, the contributions of the separate formants to age changes cannot be determined.

The results of a longitudinal study in Harrington et al. (2007) suggested that the age-dependent decrease in both F_0 and the first formant frequency might be derived from the same auditory effect of maintaining a roughly constant difference between these parameters with increasing age. More specifically, Traunmüller (1981, 1984, 1991) has shown that the difference between F_0 and F_1 in Bark is a cue for phonetic vowel height. It is therefore possible that speakers actively lower F_1 as F_0 decreases in order to maintain a roughly constant relationship between phonetic vowel height and the parameters that give rise to it perceptually with increasing age. Harrington (2006) also raises the possibility the apparent coupling of F_1 and F_0 with increasing age may be an artifact of the perturbation to formant frequencies induced by changing harmonics as F_0 decreases. Finally, as discussed above, various physiological reasons have been advanced for the decrease in formants including the possibility that the vocal tract lengthens, possibly because the larynx sinks with increasing age into the respiratory system.

In order to explore some of these ideas in further detail, we began by investigating the change in F_0 and formants in five speakers in two recordings separated by several years. Informed in part by these results, we then analysed F_0 and F_1 changes over several years within a 40–50 year period from broadcast materials of two speakers producing speech under very similar recording conditions and communicative intent. Subsequently, we made use of re-synthesis techniques to test whether the age-dependent F_1 changes could be an acoustic artifact of the decreasing F_0 and the LPC-algorithm used to calculate formants. Finally, we carried out various perception experiments to determine not only whether decreasing F_0 cues age, but also the extent to which F_1 contributes to age perception independently of F_0.

2. F_0 and formant change: A longitudinal analysis of several speakers

2.1. Method

2.1.1. Speakers and materials

This part of the analysis was based on five speakers for which longitudinal data were available. The speakers were: Queen Elizabeth II (1926-), the British actress Margaret Lockwood (1916–1990), the BBC radio presenter Roy Plomley (1914–1985), the former British prime-minister Baroness Margaret Thatcher (1925-), and the British/American journalist and broadcaster Alistair Cooke (1908–2004). The first four speakers produce a variety of the standard accent of England, Received Pronunciation; the variety spoken by Alistair Cooke is also Received Pronunciation with influences from General American.

We obtained two recordings from the BBC archives of each of these speakers which we will refer to as early and late. As shown in Table 1, the speakers' age differences vary across the two broadcasts between 29 (Lockwood) and 35 (Thatcher) years. The age of the speakers in the early broadcasts are between 34 (Queen Elizabeth II) and 43 (Cooke) years; for the late broadcasts they vary between 64 (Lockwood) and 73 (Cooke) years of age. Finally, the early broadcasts are recordings obtained between 1951 (Cooke/Plomley/Lockwood) and 1960 (Queen Elizabeth II/Thatcher); and the late broadcasts are between 1980 (Lockwood) and 1995 (Thatcher).

Table 1
The five speakers, the date of the broadcast from which the schwas were taken and the age of the speaker (in parentheses) in the early and late broadcasts. The final column shows the speakers' age difference between the early and late broadcasts.

	Early	Late	Age difference
Queen Elizabeth II	1960 (34)	1994 (68)	34
Lockwood	1951 (35)	1980 (64)	29
Thatcher	1960 (35)	1995 (70)	35
Cooke	1951 (43)	1981(73)	30
Plomley	1951 (37)	1985 (71)	34

A brief description of the corpora from which these recordings were obtained is as follows. For the Queen, the recordings were of read speech from the annual Christmas broadcasts (Harrington et al., 2000a,b,c). The data for Margaret Lockwood were from two radio presentations on a BBC program 'Desert Island Discs' in which the actress was interviewed about her life. The materials from the earlier recording were almost certainly scripted while the later recording was spontaneous speech. The data for Roy Plomley was also from 'Desert Island Discs' but not from the same broadcasts as the interviews with Lockwood. The data from Thatcher, which is spontaneous speech, was taken from a BBC radio program 'Any Questions' in which members of the panel answer questions from the audience. The recordings from Cooke were from the BBC program 'Letter from America' which was read, or rehearsed, speech.

There were considerable differences between theses five speakers in the amount of material that was available for analysis and the total durations by speaker are as follows. Queen Elizabeth II: 4 min 36 s (early), 5 min 32 s (late); Plomley: 53 s (early), 2 min 17 s (late); Thatcher 47 s (early), 1 min 16 s (late); Cooke 13 min 24 s (early), 13 min 20 s (late); Lockwood 5 min 30 s (early), 12 min (late).

2.1.2. Data pre-processing

Our interest in this study was in the long-term, non-phonetic effect of aging on pitch and vocal tract resonances and for this reason we analysed schwa vowels because we assumed that these have not been influenced by diachronic phonetic change in the last 50–60 years. Following the procedure in Harrington (2006), the schwa vowels were labelled in polysyllabic content words and care was taken not to include those weak vowels that are realized with a lax /ɪ/ in RP, i.e., schwas were only taken from weak vowels transcribed in most dictionaries with /ə/ as in *Rosa's* (but they were not taken from *roses*). Schwas were not included if their duration was so short that no reliable formant frequencies could be calculated. The sampling frequency for the speech data were 16 kHz (Queen Elizabeth II, Lockwood), 22 kHz (Cooke), and 24 kHz (Plomley, Thatcher). (The reason for this variation in the sampling frequencies is because the data had been obtained at different times and under different circumstances.) The fundamental frequency and the synchronized first four formant frequencies were calculated with a frame shift of 5 ms and a window length of 30 ms for the formant data. For

Table 2
The total number of schwas analysed in early and late broadcasts for the separate speakers.

Speaker	Broadcasts	
	Early	Late
Queen Elizabeth II	204	175
Margaret Lockwood	74	137
Margaret Thatcher	30	118
Roy Plomley	29	47
Alistair Cooke	110	140
Total	447	617

some schwas, no sensible formants could be calculated and these were discarded from further analysis. No manual correction was made to the formants of the remaining schwas. The analysis of schwas was based on F_0 and F_1–F_4 extracted at the temporal midpoint. A total of 1064 schwas were analysed, 447 from the early, and 617 from the late broadcasts (Table 2).

Because of the considerable variation in the amount of broadcast material available, there were also differences in the number of schwas available for analysis.

2.2. Results

The boxplots in Figs. 1 and 2 show that the medians and distributions of F_0 and F_1 were lower for all speakers in the late broadcasts: the possible exceptions are F_0 for Plomley and F_1 for Lockwood which show only marginal differences. We applied a linear mixed model[1] with F_0 or one of the formants as a dependent variable, Age as an independent variable, and Speaker as a random factor. The results showed significant effects for Age on both F_0 ($F[1, 60] = 98.8$, $p < 0.001$) and on F_1 ($F[1, 60] = 123.7$, $p < 0.001$).

There were no Age-dependent differences for F_2 ($F[1, 60] = 0.16$) and a weak, but not significant, trend for F_3 to increase from early to late recordings in all speakers except Lockwood for whom F_3 decreased.

The results of this longitudinal analysis are consistent with apparent-time investigations in showing a decrease in F_0 and F_1 with increasing age (e.g. Linville, 1996; Linville and Fisher, 1985a,b). In the next experiment, we investigated the relationship between decreasing F_0 and decreasing F_1 by exploring the change in these parameters within the same speakers over several years: specifically, we sought to determine the extent to which the change in F_1 was predictable from the change in F_0 or vice-versa.

3. F_0 and F_1 changes across several years in two speakers

3.1. Method

We calculated F_0 and formant frequencies in 29 Christmas broadcasts of Queen Elizabeth II and in 47 of Alistair Cooke's ‘Letter from America’ broadcasts out of 30 years. The years in which these were calculated are as follows. For Queen Elizabeth II: the 1950s (52, 54–59); 1960s (60, 62, 63–68); 1970s (70–72); 1980s (83, 85, 88); 1990s (94–99); 2000s (00–02). For Alistair Cooke, the 1940s and 1950s (47, 51, 53); the 1960s (60, 62, 65); the 1970s (70, 71, 73, 74); the 1980s (80–85); the 1990s (90–94, 96–99); and the 2000s (2000–2004). The mean duration of the 29 Christmas broadcasts was 5.2 min and of the 47 ‘Letter from America’ broadcasts 13.5 min. Because of the very large quantity of speech data that were examined (a total across all broadcasts of 2 h 35 min for the Queen and 10 h 35 min for Cooke), labelling schwas accurately following the procedure in experiment 2 would have been prohibitively time-consuming. In this experiment, we therefore opted for the faster method of calculating F_0 and F_1 averages in all voiced frames of the broadcasts. More specifically, the voiced frames were selected automatically and then concatenated into separate 1-min blocks. For each such block, the mean F_0 and mean F_1 were calculated. Finally, these means were averaged separately for each year resulting in a pair of F_0, F_1 averaged values per year per speaker.

Linear regression techniques were then applied in order to test for differences in the rate of change of F_0 and of F_1 as a function of year.

3.2. Results

Before presenting these results on F_0 and F_1 as a function of year, we first checked that voiced frames and parameters extracted at the schwa midpoint were likely to have comparable effects on age. To do this, we compared F_0 and F_1 values obtained from schwas in the preceding experiment with F_0 and F_1 values, respectively, calculated from voiced frames using the method described in Section 3.1. The results of paired sample t-tests (paired because we compared the data from schwas and voiced frames in the same year) showed no significant differences between F_0 obtained from schwas and F_0 obtained from voiced frames, neither for Queen Elizabeth II nor for Cooke; there were also no significant differences for either speaker in F_1 obtained from schwas and voiced frames. We therefore concluded that estimating the parameters from voiced frames was a viable alternative to extracting acoustic data from schwas as a means for assessing the effect of age on F_0 and F_1.

[1] One of the difficulties with mixed models is in determining the number of degrees of freedom in the denominator. An anti-conservative estimate can be obtained from $df = n - k - 1$ where n is the number of observations and k the number of degrees of freedom (Baayen, 2008). Instead of using this anti-conservative estimate (which for the data in this paper produced an estimated value of df in the denominator of between 148 and 1064), we set df to be equal to the more conservative value of 60 and chose an alpha level of 0.01. Part of the motivation for choosing a somewhat arbitrary value of 60 is that for $df > 60$ there is a fairly small change to the F-value for which significance is obtained. For example, the F values at $\alpha = 0.01$ are $F[1, 60] = 8.49$ and $F[1, 600] = 7.94$, i.e., an F-value change of 0.55 for a change in df from 60 to 600.

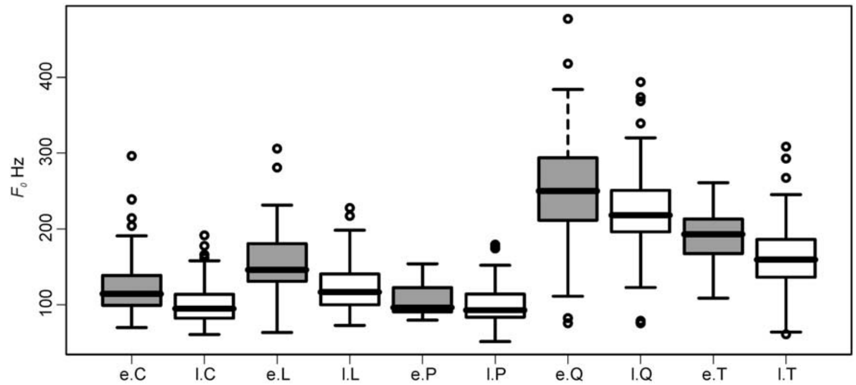

Fig. 1. Boxplots showing the distribution of F_0 in early (grey) and late (white) broadcasts extracted at the temporal midpoint of schwa for the speakers Cooke (.C), Lockwood(.L), Plomley (.P), Queen Elizabeth II (.Q) and Thatcher (.T).

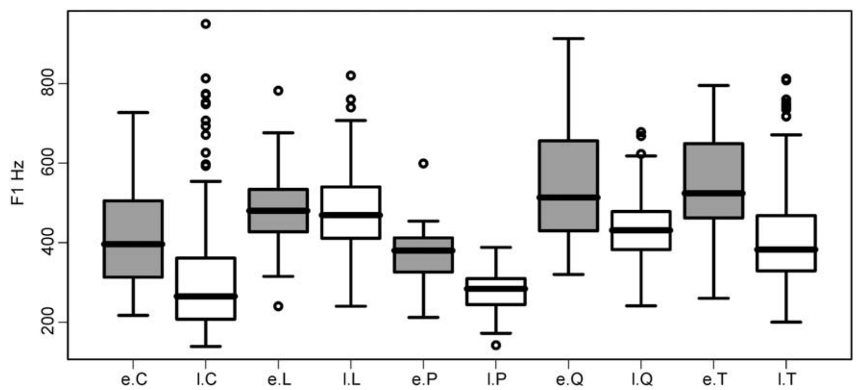

Fig. 2. Boxplots showing the distribution of F_1 in early (grey) and late (white) broadcasts extracted at the temporal midpoint of schwa for the speakers Cooke (.C), Lockwood(.L), Plomley (.P), Queen Elizabeth II (.Q) and Thatcher (.T).

We found that the most systematic relationship between F_0, F_1 and change was produced by plotting the logarithm of these parameters as a function of the year. Fig. 3 shows a plot of ln F_0, the (natural) logarithm of F_0, and ln F_1, the logarithm of F_1, as a function of speaker age in the Christmas broadcasts (left) and in the 'Letter from America' data (right). For the former, both ln F_0 and ln F_1 fall linearly with increasing age as also shown by the fitted regression lines. The data from Cooke on the right of Fig. 3 shows that both ln F_0 and ln F_1 fall up to the age of about 85,

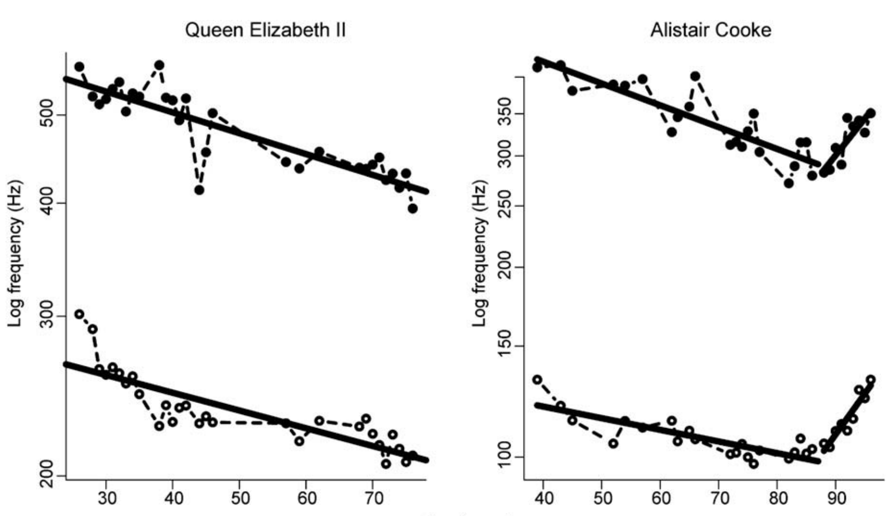

Fig. 3. Mean ln F_0 (unfilled circles) and mean ln F_1 (filled circles) in voiced frames for Queen Elizabeth II (left) and Alistair Cooke (right) as a function of chronological age with superimposed regression lines through the scatter.

and thereafter rise. Consequently, two separate regression lines were fitted to the data over the falling and rising intervals. The break-point – that is the apex of the V in this falling-rising pattern was determined algorithmically using the method of linear regression and break-point analysis discussed in detail in Baayen (2008): essentially, this method finds the year for which the sum of the squared deviations from the two regression lines is minimized. When this technique was applied to the F_0-data, the break-point was calculated to be at 87 years of age; the corresponding calculation over Cooke's F_1-data gave a break-point of 88 years of age. That these breakpoints fall within a year of each other again suggests a similar, and related trend in both F_0 and F_1 as a function of increasing age.

In order to test whether ln F_0 and ln F_1 changed at the same rate as a function of age, we compared the slopes of the regression lines using a method described in Pedhazur (1997). For the Queen, the results showed no significant differences between the regression slopes of ln F_0 and ln F_1 as a function of age (slope of F_0: −0.0045 log Hz/annum, slope of F_1 −0.0053 log Hz/annum; $F[1, 54] = 1.02$, ns). For Cooke, we compared separately the slopes of the regression lines before and after the break-point. The results showed a significantly steeper fall in F_1 than in F_0 up to the break-point age of 87 (slope of F_0: −0.0042 log Hz/annum, slope of F_1 −0.0083 log Hz/annum; $F[1, 38] = 9.45$, $p < 0.01$) and no significant difference between the rising slopes of F_0 and F_1 after the break-point (slope of F_0: 0.027 log Hz/annum, slope of F_1 0.030 log Hz/annum; $F[1, 14] = 0.16$, ns).

The general conclusion from these results is that increasing age has similar effects on F_0 and F_1: in the Queen both parameters fall, in Cooke they both exhibit a falling–rising pattern in which the change from fall to rise occurs at approximately the same age. Moreover, the rate of change in F_0 and F_1 is not significantly different except in Cooke for whom F_1 falls at a faster rate than F_0.

4. Is the relationship between F_0 and F_1 an artifact of their acoustic calculation?

Since the LPC-algorithm used for formant-estimation is founded on the principle that the source is independent of the filter, then changes in F_0 should have a negligible effect on F_1. Since on the other hand, F_1 has the greatest influence on the spectrum at the harmonics to which it is closest in frequency, then the possibility that the changing relationships between these parameters observed in the preceding section are acoustically artifactual cannot be excluded (see Harrington, 2006 for a related discussion). In order to test this possibility, we measured whether changing F_0 using synthesis techniques would have any influence on F_1.

4.1. Method

We selected for re-synthesis all schwa vowels from the 1960s broadcasts of Queen Elizabeth II (in which the Queen was 34 years of age) both because there were many schwas available in that broadcast year ($n = 200$), and because the mean F_0 for this year was approximately intermediate between the high and low fundamental frequencies of the earliest and most recent analysed broadcasts (the mean F_0s across all schwas for 1952, 1960 and 2002 were, to the nearest 5Hz, 300, 260 and 210 Hz, respectively).

A manipulation method (Moulines and Charpentier, 1990) in Praat was used to resynthesize all schwas with fundamental frequencies that were ±10% and ±20% of the original values, resulting in five tokens per schwa (the original and four resynthesized versions). Thus, the mean resynthesized F_0 extended from 310 Hz (approximately 120% of the 1960 mean of 260 Hz) to 210 Hz (−120% of the 1960 mean), i.e. a range that was comparable to the F_0 range between the 1952 and 2002 broadcasts. The formant frequencies were calculated in the original and in all F_0-resynthesized schwas using the same methodology as described in Section 2.1.2. We then ran statistical tests to determine whether any shifts in F_1 could be related to manipulated shifts in the fundamental frequency.

4.2. Results

Fig. 4 shows F_0 and F_1 averaged separately for the different synthesis conditions. Between the +20% and −10% synthesis conditions, the decreases in the averaged F_0 and averaged F_1 are comparable; however, between the −10% and −20% conditions, F_0 falls, whereas F_1 rises. We tested for trends in the data by running a repeated measures ANOVA with F_1 as the dependent variable, the five-interval synthesis condition between ±20% coded as an ordered factor, and with the schwa as a repeated factor (since five measurements were taken from the same vowel). The results showed a significant effect of synthesis condition

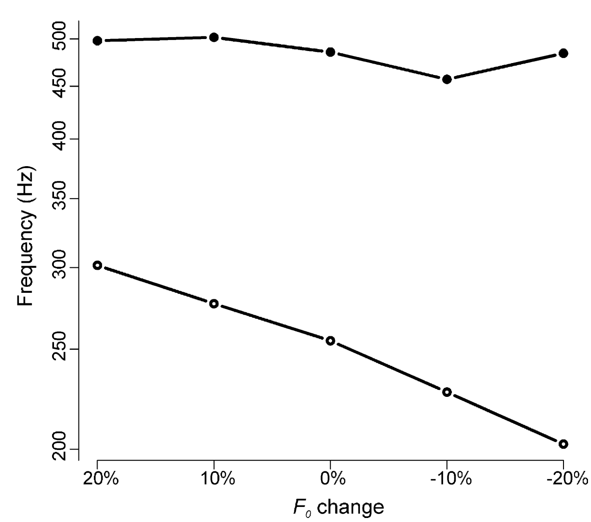

Fig. 4. Mean F_0 (unfilled circles) and mean F_1 (filled circles) in the schwas from the Christmas broadcast in 1960. The unmodified F_0 and F_1 are at 0%. The ±10% and ±20% conditions show F_0 and re-calculated F_1 values when F_0 in the schwas were changed using re-synthesis techniques to ±10% and ±20% of their original values.

on F_1 ($F[4, 796] = 22.6$, $p < 0.01$). We re-ran the statistics using a linear mixed model with schwa as a random factor to test for the types of trends. The results showed strong evidence for a linear ($t = 6.19$) and for a cubic ($t = -6.50$) trend in F_1 over the five-point synthesis interval extending between ±20%. These results are consistent with the results in Fig. 4 which, as noted earlier, show a falling trend especially between the ±10% conditions, but also a cubic shape across the entire ±20% conditions.

However, although F_0-perturbations have a significant influence on F_1, the pattern of changes is very different compared with those observed from the data in the preceding experiment: in particular, whereas the change in F_1 is similar to, or even greater than, the change in F_0 in the longitudinal data of the Queen and Cooke (Fig. 3), there is no evidence for the same pattern in the data in Fig. 4. Consequently, we conclude that the acoustic perturbation in F_1 due to a changing F_0 is not sufficient to account for the age-dependent trend in F_1 observed in the longitudinal data.

5. Perceptual judgments of age

The task in the perception experiments was to determine firstly the extent to which listeners could judge the age of the two speakers from the same, or similar, materials that had occurred in broadcasts in different years, and secondly to assess the effect of F_0 and F_1 on these judgments.

5.1. Method

Cooke introduced his ‘Letter from America’ with *good evening* for broadcasts until the mid 1980s and then *good morning* for the ones after these. We selected *good evening* from the 1947 and 1970 broadcasts and *good morning* from the 1990 broadcasts for presentation in the perceptual experiment. There were some intonational differences between the broadcasts from these 3 years, the most noticeable being that the 1947 broadcast was produced with a falling, in contrast to the low-rising, melodies in those from 1970 and 1990. The fundamental frequency and F_1 followed the pattern to be expected from Fig. 3, i.e. highest for 1947, lowest for 1970, and intermediate for 1990. The duration increased from 1947 to 1970 to 1990: this is possibly related to the decrease in duration that has been found with increasing age (Benjamin, 1982; Morris and Brown, 1987; Smith et al., 1987).

These details on Cooke are presented in Table 3 together with those from the utterances taken from the Christmas broadcasts. For the latter, we choose a longer sentence from two broadcasts that were chronologically closer together than those for Cooke. The sentence was *On earth, peace, goodwill toward men* that occurred near the end of the 1972 and also in the 1983 broadcasts, albeit with the first three words in a different order. The intonation differences between these years were confined to the first three words. As expected based on the results shown in Fig. 3, F_0 and F_1 were both higher in the 1972 than in the 1983 broadcast. The durational differences between the utterances were again confined to the first three words and came about because of the presence of an additional prosodic boundary in the 1983 materials – see Table 3 for further details.

We took some steps to reduce the differences in the signal-to-noise ratio of the broadcasts that could have indirectly provided cues to the speaker’s age. For the data from Cooke, the SNR values were 24, 29 and 37 dB for the 1947, 1970 and 1990 broadcasts, respectively. In order to reduce these differences, we scaled the mean intensity of all broadcasts and added white noise to them, such that the modified SNR values for these three broadcast years were more similar at 25, 25 and 27 dB, respectively. After applying a similar transformation to the Christmas broadcasts, the modified SNR of the 1972 and 1983 broadcasts were both 25 dB. As a further step to reducing the artifactual differences in the recordings, all five utterances were downsampled to 11 kHz.

We then derived resynthesized versions in which F_0 and F_1 were manipulated in the 1970 and 1972 broadcasts of the two speakers using TD-Psola and LPC analysis–re-synthesis

Table 3
A summary of the materials used in the perception experiment from Alistair Cooke (above) and from Queen Elizabeth II (below) showing the year of the broadcast, an intonational transcription according to the conventions of the tones-and-break indices system (round and square parentheses denote intermediate and intonational boundaries, respectively), the duration, mean F_0, and mean F_1 of the utterance. For Queen Elizabeth II the durations are those of the second intonational phrase only. For both speakers, F_0 and F_1 are averages across all voiced frames in the entire utterance.

Year	ToBI transcription	d (ms)	F_0 (Hz)	F_1 (Hz)
Alistair Cooke				
1947	H% [Good evening] L – L% L + H*	49.3	175	332
1970	[Good evening] L – H% L*	61.1	115	260
1990	H% [Good morning] L – H% L*	62.5	121	321
Queen Elizabeth II				
1972	[(on earth)H – (peace)!H – (goodwill towards men)]L – L% L + H* H* H* !H*	1260	236	430
1983	[peace on earth]L – L% [goodwill towards men]L – L% H* !H* H* !H*	1250	201	418

programs in Praat. More specifically, we obtained a total of 12 new recordings for each speaker in which F_0 and F_1 were proportionately increased and decreased in two steps either separately or together. The proportional step changes to F_0 and F_1 were ±12% and ±26%, respectively, for Cooke. Thus for Cooke, we obtained four separate resynthesized versions of the 1970 production of good evening in which only F_0 was changed by ±12% (1 step) and by ±24% (2nd step) of the original; another four separate recordings in which only F_1 was changed by ±26% and by ±52%; and finally four more recordings in which F_0 and F_1 were increased or decreased by these steps together. For Queen Elizabeth II, we also obtained 12 new recordings from her 1972 production of the sentence shown in Table 3, but with step sizes of ±20% and ±19% for F_0 and F_1, respectively. A summary of all these resynthesized data is shown in Table 4. The step sizes were estimated from F_0 and F_1 differences from schwa data that was available at the time of designing the perception experiment (and before all the data and results from the longitudinal study in experiment 3 were available). For Cooke, these steps (12% and 24%) were the percentage changes in F_0 and F_1 between the ages of 43 (1951) and 82 (1990); for Queen Elizabeth II, the corresponding percentages (20% and 19%) were the changes to F_0 and F_1 between the ages of 27 (1953) and 69 (1995). We chose those early and late steps – and thereby the big differences in F_0 and F_1 to be expected – because formant frequencies are reported to be an acoustic correlate of perceived age in whispered natural stimuli only (Linville, 1987), whereas the manipulation of speaking F_0 in (re-)synthesized stimuli is reported to contribute only very little to changes in age judgments (Winkler, 2007a,b; Harnsberger et al., 2008). So, since we were unsure of the extent of the perceptual effect of changing F_0 and F_1 by these quantities, we derived another set of materials in which the step sizes were (somewhat arbitrarily) doubled (Table 4).

Table 4
The paired entry in each cell denotes the percentage changes, respectively, to F_0 and to F_1 in deriving the 12 separate versions of Cooke's 1970 production of *good evening* and 12 versions of Queen Elizabeth II's 1972 production of *on earth*, *peace*, *goodwill towards men*.

	Cooke	Queen
F_0 only	−12, 0	−20, 0
	−24, 0	−40, 0
	12, 0	20, 0
	24, 0	40, 0
F_1 only	0, −26	0, −19
	0, −52	0, −38
	0, 26	0, 19
	0, 52	0, 38
Both F_0 and F_1	−12, −26	−20, −19
	−24, −52	−40, −38
	12, 26	20, 19
	24, 52	40, 38

There were therefore 5 unmodified stimuli (Table 3) and 24 stimuli in which F_0 and F_1 were changed separately or together (Table 4). These were repeated five times, randomized and presented to 15 listeners (thus $29 \times 5 \times 15 = 2175$ presentations). The listeners were volunteer students of phonetics at the IPS, Munich aged between 21 and 34 years (mean age 25.1 years). Thirteen of the listeners were L1-speakers of German; the other two, who had near native speaker competency in German, were L1-speakers of Hungarian and Russian. The listeners were chosen partly because of their availability, but mainly due to their unfamiliarity with the two voices to which they were asked to listen (since voice familiarity could provide indirect cues to age). They listened to the stimuli using headphones and were asked to judge the speaker's age by selecting for each stimulus a single block denoting an approximate age of the given value between the ages of 10 and 95 years (i.e., they circled one of 10, 15, 20, ... , 85, 90, 95, and were told that e.g. '35' meant 'in the middle of his/her thirties, and not necessarily exactly 35'). They were allowed as much time as they wished in listening to the stimuli and could listen to each stimulus twice.

5.2. Results

We present the results of listeners' guess of the speakers' ages in two parts: firstly in broadcasts in different years (with unmodified F_0 and F_1); and secondly in broadcasts from the same year that are identical except for modifications to F_0 and/or F_1. The first set of results (Section 5.2.1) is therefore a test of whether listeners can identify the speaker's age in stimuli for which F_0 and F_1 were unmodified; the second of the extent to which F_0 and/or F_1 contribute to judgments of the speaker's age.

5.2.1. Estimated age in different broadcasts

Fig. 5 shows the listeners' age estimates in two very similar sentences produced by Queen Elizabeth II in 1972 and 1983 and in good evening/morning produced by Cooke in 1947, 1970 and 1990. The figure suggests that listeners judged the Queen to be older in the later than the earlier broadcast and they also judged Cooke to be older in 1970 than in 1947. However, listeners judged Cooke to be of a similar, or younger, age in the 1990 than in the 1970 broadcast. We ran linear mixed models separately for each speaker with the listeners' age estimate as the dependent variable, the broadcast year as an ordered factor, and with the listener as a random factor. The results showed significant effects for the broadcast year on listeners' age estimates in both Queen Elizabeth II ($F[1, 60] = 106.8$, $p < 0.001$) and in Cooke $F[2, 60] = 40.0$, $p < 0.001$). Compatibly with Fig. 5, the mixed model analysis in Cooke showed strong evidence for both a linear ($t = 7.1$) and a quadratic ($t = -5.34$) trend.

The general conclusion from these results is that listeners correctly judged Queen Elizabeth II to be older in the second broadcast that was recorded 11 years later,

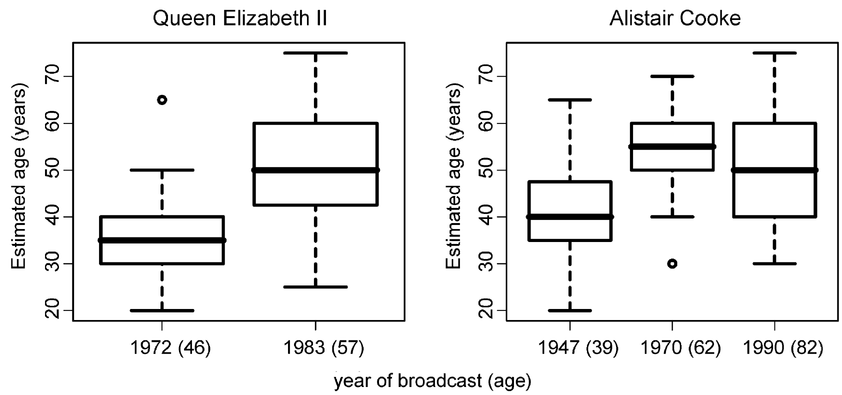

Fig. 5. The distribution of age estimated by 15 listeners in similar sentences from two Christmas broadcasts and in three similar short phrases taken from broadcasts from 'Letter from America'. The year of the broadcast and the speaker's age are shown below the boxplots.

although they did underestimate the Queen's age by some 5–10 years in both broadcasts. They also correctly judged Cooke to be older in the 1970 compared with the 1947 broadcast (and underestimated his age by roughly 5 years in the later broadcast but not in the earlier one). However, the quadratic trend in these data shows that listeners judged Cooke in 1990 to be no older, and possibly younger, than in his 1970 broadcast. The different materials (see Table 3) or perhaps even different intonational melodies could be one reason for this failure to judge the speaker to be older in the 1990 than in the 1970 broadcast. A more plausible influence is the fractionally higher mean F_0 and substantially higher F_1 in the 1990 than in the 1970 broadcast. Thus if listeners tend to associate higher F_0 or F_1 values with a lower age, then they will tend to underestimate Cooke's age beyond the break-point after which, in much older years, Cooke's F_0 and F_1 begin to increase. According to Fig. 3, the break-point is at roughly 1994/1995, whereas these data are from 1990. However, there is already evidence in Fig. 3 for some deviation from the falling trend line before the calculated break-point year, as the outlier at the age of 83 (1991 broadcast) shows; and in any case, the mean F_0 and F_1 for the particular sentence used in this perception experiment were higher than those in 1970 (Table 3).

Thus the general conclusion so far is that there is a correspondence between listeners' judgment of age and the two speakers' actual age, except for beyond the break-point in Cooke. In the next section, we tested more directly how far F_0 and F_1 contributed to this judgment.

5.2.2. *Effect of F_0 and F_1 on judgments of age*

We considered the effects of F_0 and F_1 on age estimation in two parts. Firstly, we analysed the perceptual effects of changing either F_0 or F_1 when the other parameter was unchanged. Secondly, we report on the differences between shifting F_0 alone or shifting F_0 and F_1 together.

The results for the first of these are shown separately for the two speakers in Fig. 6. The data in the left column, in which F_1 was unchanged for all F_0 manipulations, shows that perceptual estimates of age decrease dramatically in the data of Queen Elizabeth II across the five steps from lowest to highest F_0; there was a similar, but less marked, trend in the data for Cooke. The results of mixed models fitted separately to the data of the two speakers with age estimation as the dependent variable, the five F_0 steps as an ordered factor, and listener as a random factor, show that F_0 influenced age estimation significantly in the data of both the Queen ($F[4, 60] = 162.8$, $p < 0.001$) and Cooke ($F[4, 60] = 14.1$, $p < 0.001$). By contrast, the right column of Fig. 6 shows a less dramatic effect of F_1 on age estimation when F_0 is unchanged. For the Queen, the data in Fig. 6 show that a decreasing F_1 was associated with increasing age perception, an effect (using the same form of mixed model but with F_1 as the dependent variable) that was significant ($F[4, 60] = 31.3$, $p < 0.001$). On the other hand, as Fig. 6 suggests and as a statistical analysis confirms, there was no effect on age perception in Cooke's data of shifting F_1 when F_0 is unchanged.

We now consider whether F_1 contributed to age perception beyond the cues to age provided by F_0. To do this, we compared age judgments obtained from manipulating F_0 alone (and leaving F_1 unchanged) with those obtained from manipulating both F_0 and F_1 in the same direction. A summary of age estimates from these two conditions is shown in Fig. 7.

In general, Fig. 7 suggests that F_1 has a negligible effect on age judgments beyond the information provided by F_0. Thus, the leftmost two boxplots show that lowering F_0 by two steps without changing F_1 resulted in a similar estimation of age compared with lowering both F_0 and F_1 together by two steps in data from both the Queen and from Cooke. The main differences between the conditions were when the frequencies were raised. Thus, when both F_0 and F_1 were raised, the effect was to lower the estimation of age for the Queen compared with the condition in which only F_0 was raised. For Cooke, the effect was in the other direction, however: listeners judged Cooke to be older, not younger, in the condition in which F_0 and F_1 were both raised by two steps compared with the condition in which

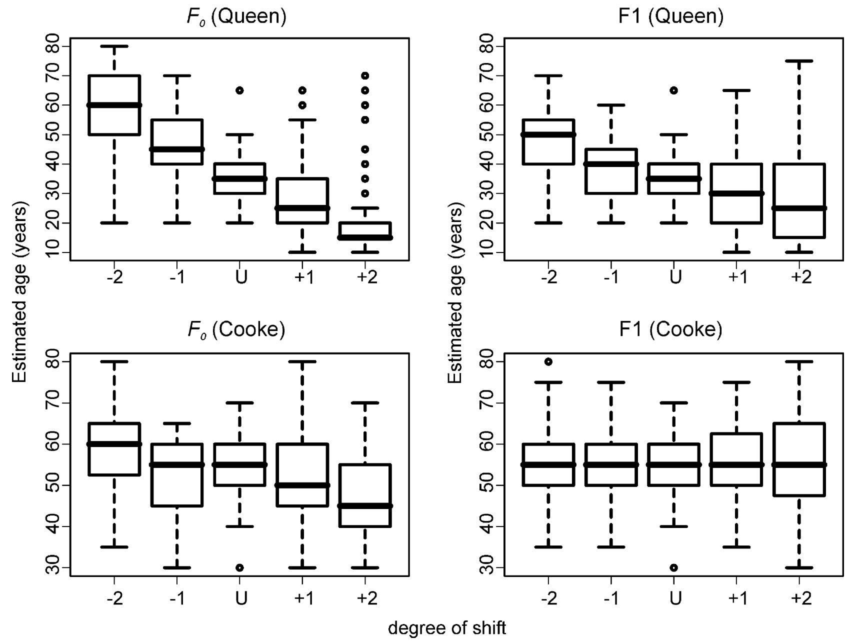

Fig. 6. The distribution of age estimated by 15 listeners in which F_0 (left column) or F_1 (right column) were resynthesized in a sentence from the 1972 Christmas broadcast (first row) and in a production of *good evening* from 'Letter from America' in 1970 (bottom row). The five conditions show the responses when the parameter was shifted in frequency downward in two steps (−2), downward in one step (−1), unchanged (U), upward by one step (+1), and upward by two steps (+2).

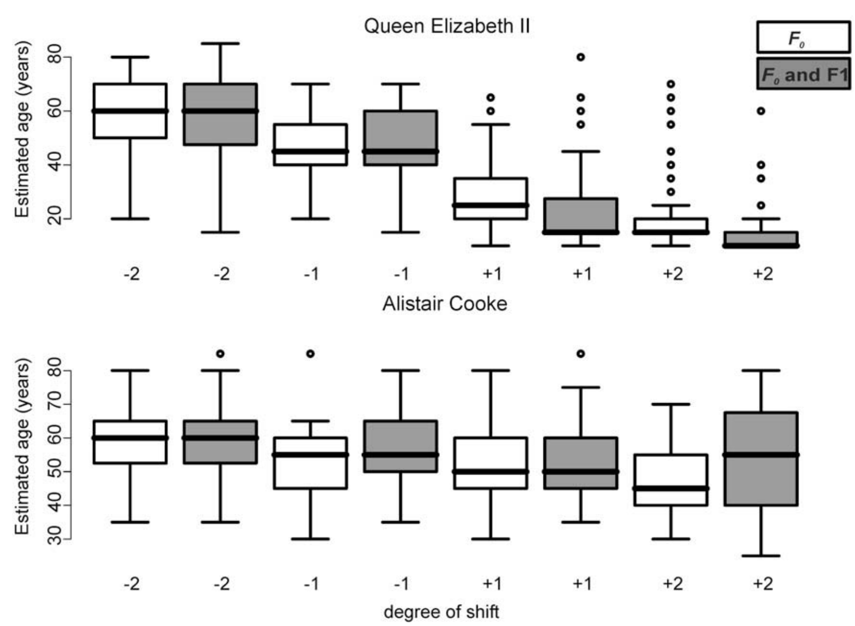

Fig. 7. The distribution of age estimated by 15 listeners in response to manipulations of just F_0 (white) or both F_0 and F_1 together (grey) in data from the 1972 the Christmas broadcasts (first row) and in a production of good evening from Letter from America in 1970 (bottom row). The four paired conditions show the responses when the parameters were shifted in frequency downward in two steps (−2), downward in one step (−1), upward by one step (+1), and upward by two steps (+2). The results shown for the grey rectangles are when both parameters are shifted by the same step size (thus −2 denotes a two-step downward shift in both F_0 and in F_1).

only F_0 was raised (far right boxplots). A subsequent mixed model analysis with dependent variable age estimate, independent factors F_0 (four levels: L2, L1, H1 and H2) and condition (two levels: manipulations of F_0 vs. manipulations of F_0 and F_1 together) and with listener as a random factor showed an effect of condition ($F[1, 60] = 9.5$, $p < 0.01$) and a significant interaction between F_0 and condition ($F[3, 60] = 4.8$, $p < 0.01$) in the data for the Queen. When we re-ran the mixed models separately for each of the four F_0 levels, the results showed a significant effect of condition only when the frequencies were raised: that is, listeners judged the Queen's age to be significantly lower when both F_0 and F_1 were raised by either one ($F[1, 60] = 22.0$, $p < 0.001$) or two ($F[1, 60] = 19.5$, $p < 0.001$) steps compared with raising F_0 on its own. For Cooke, there was a not quite significant effect of condition on perceptual age judgments: thus in general, shifts in frequency in F_0 alone or in F_0 and F_1 together produced broadly similar perceptual responses to listeners' age estimates in Cooke's data.

Summarizing across all these results, the perceptual effect of F_0 was unequivocal: raising and lowering F_0 were associated with a decrease and increase, respectively, in perceptual judgments of age. The effect of F_1 on age perception was more mixed. Although shifting F_1 while keeping F_0 unchanged resulted in corresponding shifts in perceptual age judgments, this was so only for data from the Queen, and not from Cooke. Also, there was no clear pattern of differences across both speakers when F_0 and F_1 were shifted together (upwards or downwards in frequency) compared with shifting F_0 alone. Only increasing F_0 and F_1 together in the Queen lowered age estimates more than raising F_0 on its own. Consequently, the F_0 differences were likely to have been of primary importance in listeners' perception of an older speaker in the Queen's 1983, compared with her 1972, broadcast. Also, the correct judgment for Cooke that the speaker was older in the 1970 than the 1947 broadcast, and the incorrect judgment that the speaker was younger in the 1990 than in the 1970 broadcast was most likely to be cued by mean F_0 which was highest in 1947, lowest in 1970 and intermediate between the two in 1990. The only reservation in this interpretation is that there was only a very modest F_0 increase of some 6 Hz, but a more substantial F_1 increase of 61 Hz between the 1970 and 1990 broadcasts. Thus contrary to our somewhat negative results from the second set of perception experiments showing only a very weak effect of F_1 on age perception independently of F_0, we cannot completely exclude the possibility that F_1 played some role in the listeners' (incorrect) judgments that the speaker was of a similar age, or younger, in the 1990 than in the 1970 broadcast.

6. Discussion

There were four main findings in this study. Firstly, F_0 and the first formant of schwas in at least four of the five speakers decreased between two time points sampled roughly 30 years apart. These results are consistent with those from other apparent-time (Baken, 2005; Linville, 1996; Nishio and Niimi, 2008) and longitudinal (Decoster and Debruyne, 2000; Harrington, 2007) studies in showing a decrease in F_0 with increasing age. Secondly, we found that F_0 and F_1 changed at roughly the same rate as estimated from data sampled at several intervals over a 50–60 year period in a female (Queen Elizabeth II) and a male (Alistair Cooke) speaker. Compatibly with other publications (Baken, 2005; Linville, 1996), there was clear evidence of a V-shaped trend in F_0 and a similar trend over roughly the same period in F_1 for Cooke. The results of our third study confirmed that this co-variation of F_0 and F_1 was not an artifact of the (LPC) algorithm for calculating these parameters. Finally, the results of our perception studies showed a very clear effect of F_0 on age perception, whereas, compatibly with Linville (1987) and Linville and Fisher (1985b), the perceptual effect on age of F_1 was marginal and generally not independent of F_0.

While our results are generally consistent with those from some earlier studies, there are also some differences. One of these concerns the sharp drop in F_0 that is reported by Linville (1996) for women in the age range 45–55 years that may be associated with hormonal changes due to the menopause (see also De Pinto and Hollien, 1982; Russell et al., 1995 for similar results from longitudinal studies). However, our study is more consistent with that of Baken's (2005) summary of 20 studies on F_0 and aging in showing a continuous decrease in F_0 with increasing age. Another difference is in the turning point at which F_0 in males starts to increase, after it has decreased from young adulthood: whereas Linville (1996), Baken (2005) and Brown et al. (1991) suggest turning points between the ages of 30 and 50, the trough in the age-related F_0 pattern in our male speaker was much later, and beyond 80 years of age. It may well be, then, that the year at which the trough is reached is speaker-dependent. As a Reviewer has suggested, it is possible that the late turning point may be a feature of professional speakers such as Cooke who may have learned to delay the effects of increasing age on the voice. On the other hand, professional speakers may well demonstrate even more pronounced aging effects because they use their voice so extensively: for example, a comparative study between professional and non-professional speakers by Linville (2001) showed that the effect of age on F_0 and phonation range was broadly similar in classically trained singers and non-singers when other effects such as smoking are controlled for. So the confound with the effects of other speaker variables in the apparent-time studies of Linville (1996), Baken (2005) and Brown et al. (1991) may not allow the F_0 turning point to be estimated sufficiently reliably.

We now consider some interpretations of the main finding from the acoustic part of our study that age-dependent F_1 changes seem to track those of F_0 quite closely. One possibility is that this F_0–F_1 interaction comes about because of a coupling between the laryngeal and supralaryngeal mechanisms. Thus, although the acoustic speech signal can be very well approximated by considering the source to be independent of the filter characteristics of the vocal tract, there are some well-known circumstances during which they interact. As a consequence of this interaction, F_0 can influence formants and in particular F_1 and vice-versa broadly in two ways: firstly, the shape of the glottal waveform can be skewed by vocal tract loading; and secondly, F_1 can be perturbed and dissipated by the glottis during the open phase of the glottal cycle (Childers and Wong, 1994; see also Klatt and Klatt, 1990). It is this second type of interaction which might produce a shift in F_1 of the kind observed in our results. More specifically, the open glottis impedance, which is due to a combination of the resistance and reactance produced by the mass of air in the glottal opening, can cause a small increase in F_1 frequency and quite considerable damping of F_1 (Holmes and Holmes, 2001). However, Badin and Fant's (1984) calculations based on equations by Flanagan (1965) showed the effect of glottal impedance to be small and largely irrelevant for most types of vowel production. On the other hand, in a more recent investigation of the influence of the open phase of the cycle on the acoustic response of the vocal tract using a mechanical model of vocal fold vibration, Barney et al. (2007) found that F_1 increased with increases in glottal width and the glottal opening quotient. Taking into account all of these results in the last 20–30 years, it seems that there is still some uncertainty about the magnitude of the effects on formants of this type of glottal coupling. But as far as the present data are concerned, there is in any case no evidence as far as we know that increasing age is associated with a change to the glottal width in the open phase of the glottal cycle. Moreover, even if it were possible to explain the age-dependent changes to F_1 in this way, there is still no account in this type of articulatory-to-acoustic mapping for why the frequency of F_0 and F_1 change with age at comparable rates.

Alternatively, the suggestions that the vocal tract lengthens with increasing age either because of a lowering of the respiratory and digestive system (Laver and Trudgill, 1979) or because of a lowering of the larynx in the neck (Linville and Rens, 2001) may provide a simpler account of the F_0–F_1 co-variation observed in our results. Certainly, this account is compatible with the falling F_1 observed in both the Queen and Cooke with increasing age; but at the same time, assuming the implausibility of a lengthening and then subsequent shortening of the vocal tract with increasing age, the subsequent F_1 rise in Cooke's later years requires a different explanation, possibly based on a thinning of the vocal folds, as found for older males' vocal fold epithelium (Segre, 1971), superficial (Sato and Hirano, 1997) and intermediate (Linville, 2001) layer, or on a higher incidence of glottal gap at higher age in men (Linville, 2001); both possibilities could increase the interface between sub- and supraglottal spaces, and therefore – following our first interpretation – F_1. It may also be possible to relate this F_1-change to comparable variations in F_0, given the evidence that the fundamental frequency varies in relation to the height of the larynx (Honda et al., 1999; Shipp, 1975). However, vocal tract lengthening or shortening should affect all formants, whereas the results from the present study indicate a main influence of age predominantly on F_1 (see also Harrington, 2006).

The third possibility is that the relationship between the change in F_0 and F_1 may have an auditory explanation in terms of the need to maintain a roughly constant phonetic vowel height with increasing age. Syrdal and Gopal (1986) suggested that Bark-scaled differences between F_0 and F_1 may provide a more effective cue to vowel height than F_1 alone and also that F_1–F_0 might provide an effective form of intrinsic speaker normalisation. The acoustic basis of F_1–F_0 as a cue to vowel height is the well-known finding of intrinsic pitch in which F_0 is in general higher in frequency in phonetically high than low vowels (Slawson, 1968): since F_1 decreases, i.e. changes in the opposite direction with increasing phonetic height, then the phonetic low-high vowel separation should be greater from the F_0–F_1 difference than from F_1 alone. The idea that the Bark difference between F_0 and F_1 may distinguish high from low vowels is supported by a number of perception experiments by Traunmüller (1981, 1984, 1991) showing only minimal changes to the perception of vowel height if Bark-scaled F_0 and F_1 are increased or decreased together.

Suppose now that the physiological changes to the larynx and vocal folds (see Linville, 2001 for a summary) accompanied by age-dependent changes in breathing and efficiency of laryngeal airway valving (Hoit and Hixon, 1987, 1992; Huber and Spruill, 2008; Melcon et al., 1989) bring about the kinds of changes to fundamental frequency that have been observed in this paper. Unless F_1 changes as well, then, according to this auditory theory, the cues to phonetic vowel height will be disrupted. Consequently, it is possible that speakers actively lower F_1 in response to their age-dependent changes in F_0 in order to maintain a relatively constant perception of phonetic vowel height with increasing age. Such an F_1-adjustment could be made by changing the extent of the mouth opening. As Lindblom and Sundberg (1971) show in their predictions of resonances from models of the vocal tract, the most consistent effect of jaw lowering (and hence mouth opening) is an F_1-raising for all types of constriction location and positions. Moreover, when the mouth opening was increased in a single tube model of the vocal tract for schwa in their study, there was no effect on F_2, while the rise in F_3 was much more modest than that for F_1 (and for jaw openings beyond 10 mm, F_3 did not rise at all). These predicted formant adjustments in response to adjustment of the mouth opening fit our observations on age-related changes quite well. Thus, one plausible interpretation of our data is that the physiological changes to the larynx and vocal folds that produce a falling F_0 with increasing age (and then a rising F_0 in later years in male speakers) are offset by small adjustments to the mouth opening in order to maintain a roughly constant auditory distance between F_0 and F_1, thereby minimizing changes to perceived vowel height with increasing age. This interpretation is consistent with a recent study by Beyerlein et al. (2008) in which an acoustic-to-articulatory inversion was applied using Maeda's model (Maeda, 1979, 1990) to the same longitudinal data of Queen Elizabeth II: their results showed a progressively raised jaw position and therefore a smaller mouth opening in the Queen's data with increasing age.

Finally, our results suggest the need for caution in the interpretation in apparent-time studies of differences in F_1 between younger and older speakers as indicative of phonetic sound change. Thus a decrease in F_1 due to diachronic raising may be confounded with the lower F_1 that results from age differences. One possible way of disentangling this confound might be to take the speaker's F_0 into account and base the diachronic change to phonetic height on the auditory distance between F_1 and F_0. Alternatively, and in our view much more reliably, this confound can be overcome as in Harrington et al. (2008) by assessing separately for each speaker the position in a formant space of a vowel subject to a sound change in progress relative to another vowel, or set of vowels, that are comparatively diachronically stable.

References

Abitbol, J., Abitbol, P., Abitbol, B., 1999. Sex hormones and the female voice. Journal of Voice 13, 424–446.

Baayen, R.H., 2008. Analyzing Linguistic Data. A Practical Introduction to Statistics Using R. Cambridge University Press.

Badin, P., Fant, G., 1984. Notes on vocal tract computation. Speech Transmission Laboratory, Quarterly Progress Status Report, 2–3, pp. 53–108.

Bailey, G., Wikle, T., Tillery, J., Sand, L., 1991. The apparent time construct. Language Variation and Change 3, 241–264.

Baken, R.J., 2005. The aged voice: a new hypothesis. Journal of Voice 19 (3), 317–325.

Barney, A., De Stefano, A., Henrich, N., 2007. The effect of glottal opening on the acoustic response of the vocal tract. Acta Acustica 93, 1046–1056.

Beckman, M.E., Venditti, J.J., 2010. Tone and intonation. In: Hardcastle, W.J., Gibbon, F., Laver, J. (Eds.), Handbook of Phonetic Sciences, second ed. Blackwell, pp. 603–652.

Benjamin, B., 1982. Phonological performance in gerontological speech. Journal of Psycholinguistic Research 11, 159–167.

Beyerlein, P., Cassidy, A., Kholhatkar, V., Lasarcyk, E., Nöth, E., Potard, B., Shum, S., Song, Y.C., Spiegl, W., Stemmer, G., Xu, P., 2008. Vocal Aging Explained by Vocal Tract Modelling: 2008 JHU Summer Workshop Final Report. <http://www.clsp.jhu.edu/workshops/ws08/documents/ws08vaeFinalReport.pdf>.

Brown Jr., W.S., Morris, R.J., Hollien, H., Howell, E., 1991. Speaking fundamental frequency characteristics as a function of age and professional singing. Journal of Voice 5 (4), 310–315.

Campbell, N., 2004. Databases of expressive speech. Journal of Chinese Language and Computing 14 (4), 295–304.

Chambers, J.K., Trudgill, P., Schilling-Estes, N. (Eds.), 2002. The Handbook of Language Variation and Change. Blackwell Publishers.

Childers, D., Wong, C., 1994. Measuring and modeling vocal source–tract interaction. IEEE Transactions on Biomedical Engineering 41, 663–671.

Cox, F., 2006. The acoustic characteristics of /*hVd*/ vowels in the speech of some Australian teenagers. Australian Journal of Linguistics 26, 147–179.

Decoster, W., Debruyne, F., 2000. Longitudinal voice changes: facts and interpretation. Journal of Voice 14, 184–193.

De Jong, K.J., 1995. The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation. Journal of the Acoustical Society of America 97, 491–504.

De Pinto, O., Hollien, H., 1982. Speaking fundamental frequency characteristics of Australian women: then and now. Journal of Phonetics 10, 367–375.

Docherty, G., in press. Speaker-related variation – sociophonetic factors. In: Cohn, A.C., Fougeron, C., Huffman, M., (Eds.), The Oxford Handbook of Laboratory Phonology. Oxford University Press, Oxford.

Endres, W., Bambach, W., Flosser, G., 1971. Voice spectrograms as a function of age, voice disguise, and voice imitation. Journal of the Acoustical Society of America 49, 1842–1847.

Flanagan, J.L., 1965. Speech Analysis, Synthesis, and Perception. Springer-Verlag, New York.

Flügel, C., Rohen, J.W., 1991. The craniofacial proportions and laryngeal position in monkeys and man of different ages (A morphometric study based on CT-scans and radiographs). Mechanisms of Ageing and Development 61, 65–83.

Foulkes, P., Docherty, G.J., 2006. The social life of phonetics and phonology. Journal of Phonetics 34 (4), 409–438.

Gugatschka, M., Kiesler, K., Obermayer-Pietsch, B., Schoekler, B., Schmid, C., Groselj-Strele, A., et al., in press. Sex hormones and the elderly male voice. Journal of Voice (corrected proof, available online).

Harnsberger, J.D., Shrivastav, R., Brown, W.S., Rothman, H., Hollien, H., 2008. Speaking rate and fundamental frequency as speech cues to perceived age. Journal of Voice 22 (1), 58–69.

Harrington, J., 2006. An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. Journal of Phonetics 34, 439–457.

Harrington, J., 2007. Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. In: Laboratory Phonology, vol. 9. Mouton, Berlin.

Harrington, J., Fletcher, J., Beckman, M.E., 2000a. Manner and place conflicts in the articulation of accent in Australian English. In: Broe, M. (Ed.), . In: Papers in Laboratory Phonology, vol. 5. Cambridge University Press, Cambridge, pp. 40–55.

Harrington, J., Palethorpe, S., Watson, C., 2000b. Does the Queen speak the Queen's English?. Nature 408 927–928.

Harrington, J., Palethorpe, S., Watson, C., 2000c. Monophthongal vowel changes in Received Pronunciation: an acoustic analysis of the Queen's Christmas broadcasts. Journal of the International Phonetic Association, 63–78.

Harrington, J., Palethorpe, S., Watson, C., 2005. Deepening or lessening the divide between diphthongs? An analysis of the Queen's annual Christmas broadcasts. In: Hardcastle, W.J., Beck, J. (Eds.), The Gift of Speech (Festschrift for John Laver). Erlbaum, pp. 227–261.

Harrington, J., Palethorpe, S., Watson, C.I., 2007. Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers. In: INTERSPEECH-2007, Antwerp, pp. 2753–2756.

Harrington, J., Kleber, F., Reubold, U., 2008. Compensation for coarticulation, /u/-fronting, and sound change in Standard Southern British: an acoustic and perceptual study. Journal of the Acoustical Society of America 123, 2825–2835.

Hoit, J.D., Hixon, T.J., 1987. Age and speech breathing. Journal of Speech and Hearing Research 30, 351–366.

Hoit, J.D., Hixon, T.J., 1992. Age and laryngeal airway resistance during vowel production in women. Journal of Speech and Hearing Research 35 (2), 309–313.

Hollien, H., Shipp, T., 1972. Speaking fundamental frequency and chronologic age in males. Journal of Speech and Hearing Research 15, 155–159.

Holmes, J., Holmes, W., 2001. Speech Synthesis and Recognition, second ed. Taylor & Francis, London.

Honda, K., Hirai, H., Masaki, S., Shimada, Y., 1999. Role of vertical larynx movement and cervical lordosis in F_0 control. Language and Speech 42 (4), 401–411.

Huber, J.E., Spruill III, J., 2008. Age-related changes to speech breathing with increased vocal loudness. Journal of Speech, Language, and Hearing Research 51 (3), 651–668.

Klatt, D.H., Klatt, L.C., 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. Journal of the Acoustical Society of America 87 (2), 820–857.

Labov, W., 1972. Sociolinguistic Patterns. University of Pennsylvania Press, Philadelphia.

Labov, W., 1994. Internal factors. In: Principles of Linguistic Changes, vol. 1. Blackwell, Oxford.

Labov, W., 2001. Social factors. In: Principles of Linguistic Changes, vol. 2. Blackwell, Oxford.

Laver, J., Trudgill, K., 1979. Phonetic and linguistic markers in speech. In: Scherer, K.R., Giles, H. (Eds.), Social Markers in Speech (European Studies in Social Psychology). Cambridge Univ. Press, Cambridge.

Lindblom, B., Sundberg, J., 1971. Acoustical consequences of lip, tongue, jaw, and larynx movement. Journal of the Acoustical Society of America 50, 1166–1179.

Linville, S.E., Fisher, H., 1985a. Acoustic characteristics of women's voices with advancing age. Journal of Gerontology 40 (3), 324–330.

Linville, S.E., Fisher, H., 1985b. Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females. Journal of the Acoustical Society of America 78, 40–48.

Linville, S.E., 1987. Acoustic-perceptual studies of aging voice in women. Journal of Voice 1, 44–48.

Linville, S.E., 1996. The sound of senescence. Journal of Voice 10, 190–200.

Linville, S.E., Rens, J., 2001. Vocal tract resonance analysis of aging voice using long-term average spectra. Journal of Voice 15 (3), 323–330.

Linville, S.E., 2001. Vocal Aging. Singular Publishing Group, San Diego.

Maeda, S., 1979. Un modèle articulatoire de la langue avec des composantes linéaires. In: Actes 10èmes Journées d'Etude sur la Parole, Grenoble, pp. 152–162.

Maeda, S., 1990. Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In: Hardcastle, W., Marchal, A. (Eds.), Speech Production and Speech Modelling. Kluwer Academic Publisher, Amsterdam, pp. 131–149.

Melcon, M., Hoit, J., Hixon, T., 1989. Age and laryngeal airway resistance during vowel production. Journal of Speech and Hearing Disorders 54, 282–286.

Mendoza-Denton, N., in press. Inter- and Intra-speaker Variation in LabPhon and Field-based Sociophonetics. Oxford University Press, Oxford.

Morris, R., Brown, W., 1987. Age-related voice measures among adult women. Journal of Voice 1, 38–43.

Moulines, E., Charpentier, F., 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. Speech Communication 9, 453–467.

Mwangi, S., Spiegl, W., Hönig, F., Haderlein, T., Maier, A., Nöth, E., 2009. Effects of vocal aging on fundamental frequency and formants. In: Acoustical Society of the Netherlands (NAG); German Acoustical Society (DEGA) (Eds.), Proceedings of the International Conference on Acoustics NAG/DAGA 2009, pp. 1761–1764.

Nishio, M., Niimi, S., 2008. Changes in speaking fundamental frequency characteristics with aging. Folia Phoniatrica et Logopaedica 60, 120–127.

Pedhazur, E.J., 1997. Multiple Regression in Behavioral Research, third ed. Harcourt Brace, Orlando, FL.

Rastatter, M., McGuire, R., Kalinowski, J., Stuart, A., 1997. Formant frequency characteristics of elderly speakers in contextual speech. Folia Phoniatrica et Logopaedica 49, 1–8.

Russell, A., Penny, L., Pemberton, C., 1995. Speaking fundamental frequency changes over time in women: a longitudinal study. Journal of Speech and Hearing Research 38, 101–109.

Sankoff, G., Blondeau, H., 2007. Language change across the lifespan: /r/ in Montreal French. Language 83 (3), 560–588.

Sato, K., Hirano, M., 1997. Age-related changes of elastic fibers in the superficial layer of the lamina propria of vocal folds. Annals of Otology, Rhinology and Laryngology 106, 44–48.

Segre, R., 1971. Senescence of the voice. Eye, Ear, Nose & Throat Monthly 50, 62–68.

Scukanec, G., Petrosino, L., Squibb, K., 1991. Formant frequency characteristics of children, young adult, and aged female speakers. Perceptual and Motor Skills 73, 203–208.

Shipp, T., 1975. Vertical laryngeal position during continuous and discrete vocal frequency change. Journal of Speech and Hearing Research 18, 707–718.

Slawson, A.W., 1968. Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. Journal of the Acoustical Society of America 43 (1), 87–101.

Smith, A., Wasowicz, J., Preston, J., 1987. Temporal characteristics of the speech of normal elderly adults. Journal of Speech and Hearing Research 30, 522–529.

Syrdal, A.K., Gopal, H.S., 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. Journal of the Acoustical Society of America 79 (4), 1086–1100.

Traunmüller, H., 1981. Perceptual dimension of openness in vowels. Journal of the Acoustical Society of America 69, 1465–1475.

Traunmüller, H., 1984. Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels. Speech Communication 3, 49–61.

Traunmüller, H., 1991. The context sensitivity of the perceptual interaction between F_0 and F_1. In: Proceedings of the 12th International Congress of Phonetic Sciences, vol. 5, Aix-en-Provence, pp. 62–65.

Verdonck-de Leeuw, I.M., Mahieu, H.F., 2004. Vocal aging and the impact on daily life: a longitudinal study. Journal of Voice 18 (2), 193–202.

Weinreich, U., Labov, W., Herzog, M., 1968. Empirical foundations for a theory of language change. In: Lehmann, W.P., Malkiel, Y. (Eds.), Directions for Historical Linguistics. University of Texas Press, Austin, TX.

Wind, J., 1970. On the Phylogeny and Ontogeny of the Human Larynx. Wolters-Noordhoff Publishing, Groningen.

Winkler, R., 2007a. Influences of pitch and speech rate on the perception of age from voice. In: Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken.

Winkler, R., 2007b. Testing the relevance of speech rate, pitch and a glottal chink for the perception of age in synthesized speech using formant synthesis. In: Proceedings of Interspeech 2007, Antwerp.

Xue, S.A., Hao, G.J., 2003. Changes in the human vocal tract due to aging and the acoustic correlates of speech production: a pilot study. Journal of Speech, Language, and Hearing Research 46, 689–701.

Zemlin, W.R., 1998. Speech and Hearing Science – Anatomy and Physiology, fourth ed. Allyn and Bacon, Boston.