Twenty-eight years of vowels: Tracking phonetic variation through young to middle age adulthood

Susanne Gahl*, Harald Baayen

UC Berkeley, USA

Eberhard Karls University Tübingen, Germany

Abstract

Research on age-related changes in speech has primarily focused on comparing "young" vs. "elderly" adults. Yet, listeners are able to guess talker age more accurately than a binary distinction would imply, suggesting that acoustic characteristics of speech change continually and gradually throughout adulthood. We describe acoustic properties of vowels produced by eleven talkers based on naturalistic speech samples spanning a period of 28 years, from ages 21 to 49. We find that the position of vowels in F1/F2 space shifts towards the periphery with increasing talker age. Based on Generalized Additive Mixed-effects regression models, we show that this shift is not fully attributable to changes in vowel duration or to segmental context. We discuss the implications of our results for research on aging and speech, and for research in which durational shortening and spectral characteristics of vowels are assumed to reflect a unitary process of phonetic reduction.

Keywords: phonetic reduction, vowels, vowel duration, vowel formants,

*Preprint, published with minor changes: Gahl, S. & R.H. Baayen. 2019. Journal of Phonetics 74.

Email address: gahl@berkeley.edu (Susanne Gahl)

talker age, Generalized Additive Models

1. Introduction

Articulatory and acoustic properties of vowels depend in part on vowel duration: Shortening may be associated with articulatory undershoot, resulting in a reduced range of the first two vowel formants (Lindblom, 1963, 1964; Fourakis, 1991), whereas lengthening is often associated with greater articulatory movement (Smiljanić and Bradlow, 2009; Bradlow, 2002; Moon and Lindblom, 1994). And yet, durational shortening does not inevitably result in reduced articulatory movements or vowel space contraction (see e.g. Lindblom, 1983; Gay, 1978; Bradlow, 2002; Clopper et al., 2017). Conversely, vowel lengthening does not necessarily entail vowel space expansion (see e.g. Fletcher et al., 2015). Despite the partial independence of vowel duration and articulation, there is a commonly used research strategy which treats changes in duration, vowel space size, and articulatory range as different manifestations of a single gradient phenomenon, commonly referred to as *phonetic reduction* and *phonetic enhancement* (or *strengthening*). If duration (of segments, syllables, and words), vowel formants, and other acoustic and articulatory parameters are manifestations of a single phenomenon, then they make equally suitable variables in investigations of the causes and consequences of phonetic variation. Numerous studies, including some of our own, have adopted that strategy and used a single variable as an index of phonetic reduction (for examples and overviews see e.g. Bell et al., 2009; Gahl, 2008; Gahl and Strand, 2016; Seyfarth, 2014; Fink and Goldrick, 2015; Jaeger and Buz, 2016). For many practical purposes, using one variable as a proxy for phonetic reduction generally is not only convenient, but also reasonable: Studies considering multiple dimensions of 'reduction' have often found multiple variables to converge (e.g. Aylett and Turk 2006; Gahl and Garnsey 2004; Gahl et al. 2012; Harnsberger et al. 2008; Brink et al. 1998; Son et al. 2003; Son and Pols 2003).

However, the partial independence of different phonetic parameters such as vowel formants and vowel duration makes the assumption of across-the-board reduction problematic. A set of recent studies highlights this issue. Tomaschek et al. (2013, 2014, 2018b) found that vowels in high-frequency words were shorter, but more peripheral in F1/F2 space than vowels in low-frequency words. Tomaschek et al. (2014) relate their findings to *learning*: "Higher frequency not only enables the speaker to articulate more efficiently but also more precisely." That conclusion highlights the potential role of learning and usage experience in articulation, duration, and acoustic characteristics of vowels. Viewed in this way, estimates of word frequency are an estimate of individual talkers' linguistic experience, i.e. how often a speaker may have heard or spoken a given word.

Another measure of the amount of usage experience a talker has had is the talker's age, the focus of the current study. We hypothesized that pronunciation would change over the adult life span in a manner analogous to the changes observed for high-frequency words in Tomaschek et al. (2013). If usage frequency and talker age indeed have partially analogous effects, then talkers' overall vowel spaces should expand as talkers move through young-to-middle age adulthood, as a result of maximally peripheral targets increasingly being realized along the periphery even at short durations.

To test our hypotheses, we analyzed vowel formants (F1 and F2) in a corpus of speech samples of eleven talkers recorded once every seven years, starting at age 21 and ending at age 49.

2. Background

2.1. Age-related changes in vowel space size and/or the first two formants

The evidence on the development of vowel spaces during young to middle-age adulthood is limited. Moreover, many studies using between-group designs involve groups that are either narrow, but spaced fairly far apart (e.g. comparing age 20-25 vs. 45-50 vs. 70-75) or else broad and close together (e.g. 30-50 vs. 50 and up). Both of these designs get in the way of detecting developments during middle age adulthood. Figure 1 shows the age groups covered in between-group studies of F1 and F2, based on a survey of the literature; the survey was initially based on references cited in Baken and Orlikoff (2000) and Linville (2001), followed by searches for 'talker age', 'vowels', and 'vowel formants' in PsycInfo and Google scholar, conducted at several points in the past 5 years, and on the citation trail of those references forward and backward in time. The picture that emerges from the admittedly unsystematic survey at this point is that the majority of studies compare adults in their twenties vs. 70 years and up. These comparisons are nevertheless relevant to the present study: As we shall see shortly, there are several patterns of differences in vowel spaces of young vs. elderly adults that have been noted with some consistency. Given how few studies have tracked development within young and middle-age adulthood, these differences may conceivably begin to emerge during middle-age adulthood. Alternatively, changes over the adult lifespan may unfold in a non-linear fashion: For example, vowel spaces might expand during middle age and then contract at some later point.

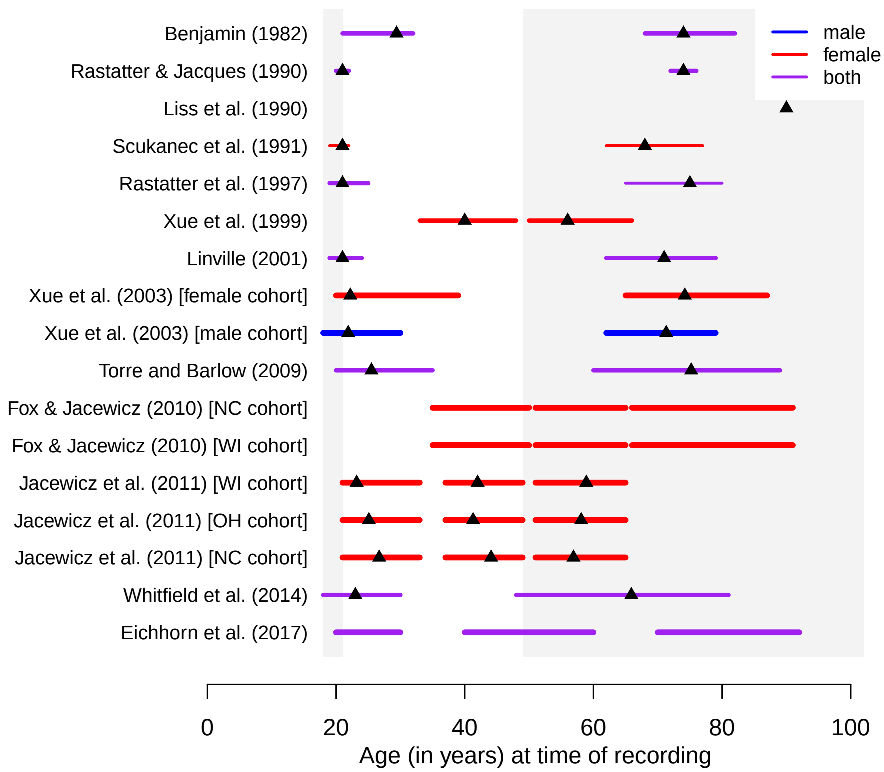

Figure 1: Cross-sectional studies of effects of age on F1 and F2 in adult talkers. The unshaded area from age 21 to 49 indicates the age span in our corpus. Horizontal bars represent the age range (if reported) of each group of talkers included. Triangles represent the mean age (if reported) of each group of talkers. Horizontal bars on the same row show groups or individuals compared to one another in a single study. Color coding is used to indicate the sex of the participants.

As our hypothesis concerns vowel space size, we first ask what is known about the relationship between talker age and vowel centralization, i.e. contraction of the F1/F2 space, or expansion, i.e. movement towards the periphery), and about vowel positions within F1/F2 space. Several studies comparing adults in their twenties and thirties vs. age 70 and up have found that vowels centralize as talkers get older. Evidence for centralization has taken three different forms: changes in vowel space size (Fox and Jacewicz, 2010), reduction in F1 and/or F2 range (Benjamin, 1982; Xue and Hao, 2003; Rastatter and Jacques, 1990; Rastatter et al., 1997; Torre and Barlow, 2009) in older vs. younger adults, and reduced distance between peripheral vowels and schwa (Liss et al., 1990) in elderly adults compared to published reference values in younger adults (Stevens et al., 1966).

Changes in vowel space sizes may take place concurrently with changes in the positions of individual vowels or across-the-board shifts in vowel space position. For F1, there is some evidence of overall shifts, specifically decreases in F1 of both high and low vowels (Linville, 2001; Torre and Barlow, 2009; Jacewicz et al., 2011; Xue and Hao, 2003). A decrease in F1 for the vowels [i, u, æ] and [ɑ] was also observed by Scukanec et al. (1991), along with F2 decreasing in the back vowels [u] and [ɑ]. This pattern suggests a general raising of the space, possibly in combination with an expansion of the space along the F2 dimension (though overall size of vowel spaces was not reported in Scukanec et al. 1991). Similarly, patterns of change noted in Schötz and Müller (2007); Price (2006) point to F1 decreases and concurrent vowel space expansion.

Few studies include comparisons within the 21-to-49 year age range or close to it. Studies that do report such comparisons do report systematic changes, however, suggesting that vowel spaces are indeed not entirely stable in this age group. The observed directions of these changes are consistent with overall changes in vowel space position (e.g. overall raising and/or fronting), changes in vowel space size (e.g. due to decreasing or increasing F2 range), or both: Jacewicz et al. (2011) found evidence of F1 lowering with increasing talker age for [i,ɪ,æ], whereas Xue et al. (1999) found that F2 in [a] was higher in a group of older female talkers (mean age = 56) than in the younger one (mean age = 40). Evidence for expansion of vowel space size is presented in Eichhorn et al. (2018); while vowel space sizes are not reported numerically, visual inspection and measurements based on Figure 3 in Eichhorn et al. (2018) suggest that vowel spaces were in fact larger for the groups aged 40-60 years, compared to the groups aged 20-30 years, especially for the female participants.

Complementing between-group designs, several studies have used speech corpora unselected for age: These studies offer the potential for tracking non-linear changes across large age ranges (see e.g. Schötz and Müller, 2007 for analyses of a wide range of acoustic parameters in Swedish, and Horton et al., 2010 for an analysis of speaking tempo in the Switchboard corpus of American English). Two such studies (Hay et al., 2015; Fletcher et al., 2015) track the development of F1 and F2 in speakers of English. Hay et al. (2015) examined F1 and F2 in a corpus of New Zealand English (the ONZE corpus, Gordon et al., 2004). Unfortunately, as Hay et al. (2015) point out, "[t]he earlier born speakers were older when they were recorded than the later born speakers" (Hay et al., 2015, p.86) in that corpus, making it difficult to tease apart age-related changes in individual talkers from dialect changes. Fletcher et al. (2015) examined the vowel spaces (defined as the area of the triangle formed by the vowels [ɐː,iː,oː]) in read speech by 149 speakers of New Zealand English aged 65-90 (i.e. roughly forty years older than the ones in our corpus). Two findings emerged: The first is that talkers who habitually spoke more slowly tended to have larger vowel spaces than talkers who spoke more quickly. The second finding is that speaking tempo decreased with increasing talker age without a concomitant change in vowel space size. Both of these findings underscore the need to control for vowel duration in studies of vowel position in F1/F2 space, including the current study.

To our knowledge, longitudinal studies of vowel space sizes during the period spanning ages 21 through 49 have so far mostly been restricted to samples of read and highly scripted speech of public figures. Possibly indicating an increase in vowel space size during the period most relevant to the current study Harrington (2006), in an analysis of Queen Elizabeth's Christmas address, found that the F1 range increased during the Queen's young to middle age adulthood. Given the differences between read speech and spontaneous speech, it is not clear whether one should expect a similar pattern in the unscripted monologues and responses to interview questions in our corpus.

2.2. Fundamental frequency and concomitant changes in F1

Age-related changes in f0 have been argued to be coupled to changes in F1 (Reubold et al., 2010); therefore, studies reporting f0 may indirectly offer some clues to changes in F1.

Numerous studies provide relevant information, both based on cross-sectional designs (see e.g. Cox and Selent, 2015; Goy et al., 2013; Hollien et al., 1997; Morris and Brown, 1994; Ramig, 1986; Selent, 2014; Stathopoulos et al., 2011; Torre and Barlow, 2009; Xue and Deliyski, 2001) and longitudinal ones (e.g. Verdonck-de Leeuw and Mahieu, 2004; Decoster and Debruyne, 2000; Reubold et al., 2010; Harrington et al., 2007; Quené, 2013). Most observations about f0 are consistent with a pattern in which f0 decreases for both male and female talkers from young adulthood until some point somewhere between ages 50 and 65 (see Baken, 2005 for an overview, as well as Cox and Selent, 2015; Sataloff et al., 1997; Hollien and Shipp, 1972; Bier et al., 2017, among many others), though there are exceptions to this pattern (see e.g.Braun and Friebis, 2009; Nishio and Niimi, 2008). In fact, some studies (e.g. Linville, 1996) report a fairly sharp drop in women's f0 as young as 45-55 years, possibly associated with hormonal changes. After age 50 (or later), women's and men's f0 appear to follow different trajectories, increasing in men and decreasing in women. If age-related changes in F1 track changes in f0, then these observations are consistent with the evidence reviewed above, suggesting F1 decreases during the same period. As for longitudinal studies, Reubold et al. (2010) and Reubold and Harrington (2015) analyzed the average values of f0 and F1 in recordings of the speech of Queen Elizabeth II and journalist Alistair Cooke spanning over fifty years. In both of these speakers, the average f0 and F1 values observed at the earliest recording (age 26 for Queen Elizabeth, age 39 for Alistair Cooke) are higher than at the oldest age before age 50 (age 46 for the Queen, age 45 for Cooke), consistent with an overall decrease in f0 and F1 during the period spanned in our corpus.

2.3. Temporal measures

Given that longer vowels tend to be more peripheral in F1/F2 than shorter ones, then any change in vowel space we observe could be the result of changes in speaking rate. To check whether that scenario is likely, we must consider age-related changes in overall speaking tempo, vowel durations, and articulatory velocity.

Figure 2 shows the coverage of cross-sectional studies of speaking tempo. The studies including comparisons in the age range of interest here provide very little evidence for age-related lengthening of segment or syllable durations in young-to-middle age speakers, and some evidence to the contrary. As for individual articulatory movements, Dromey and Scott (2016) did not observe any differences across age-groups in the duration of a target sentence or in speech kinematic measures, such as upper and lower lip displacement and velocity.

Intriguingly, and most relevantly for the current study, Jacewicz et al. (2010) found a pattern consistent with the notion that speaking tempo may in fact increase during young-to-middle age adulthood. Jacewicz et al. (2010) found a non-linear (inverted U-shaped) pattern, with speaking tempo in spontaneous speech (measured as syllables per second) increasing until about speakers' late forties and decreasing after that. Jacewicz et al. (2010) further found that speaking rate for reading aloud was fastest for the younger group of speakers aged 20-34. Jacewicz et al. (2010) points out that the differences between spontaneous speech (where older adults were faster than young adults) and reading rate may be due to the fact that college-aged adults read extensively on a daily basis, perhaps more than the elderly adults enrolled

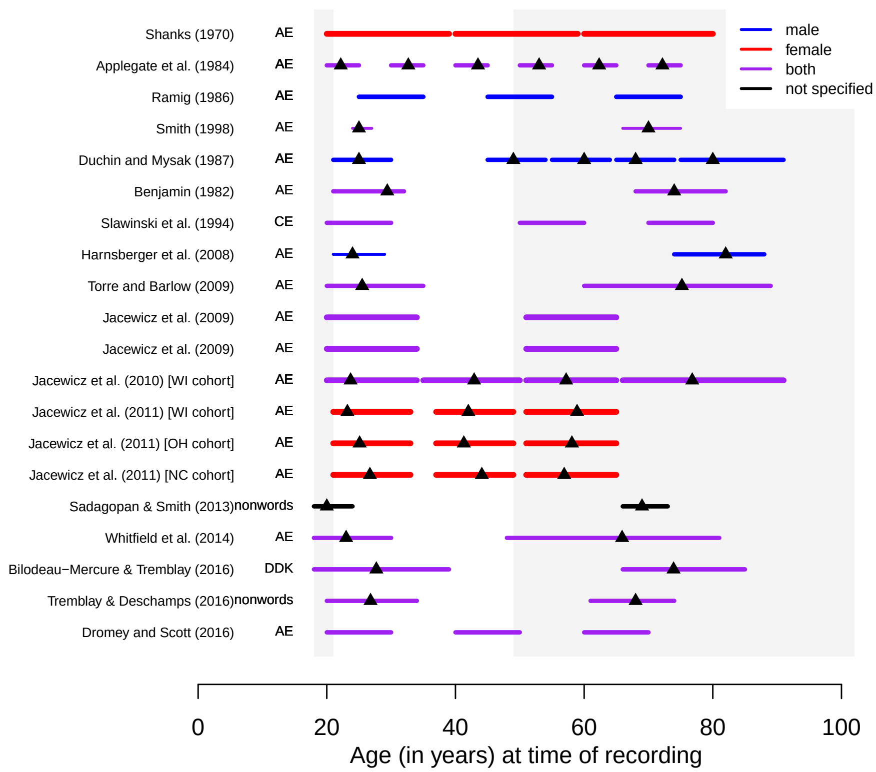

Figure 2: Cross-sectional studies of age-related changes in temporal characteristics of speech in English and using diadochokinetic rates (see text). AE = US American English, CE = Canadian English, DDK = diadochokinetic rate. The unshaded area from age 21 to 49 indicates the age span in our corpus. Triangles represent the mean age (if reported) of each group of talkers. Horizontal bars represent the age range (if reported) of each group of talkers included. Horizontal bars on the same row show groups or individuals compared to one another in a single study. Color coding is used to indicate the sex of the participants, where reported.

in the study.

One of a very small number of studies to examine changes in speaking tempo longitudinally, Quené (2013) investigated speaking tempo in a set of nine speeches recorded during ages 42 through 74 in the life of Queen Beatrix of the Netherlands. Overall articulation rate was found to decrease from age 42 up to about age 58, and then to increase. Within each speech, there was a clear pattern of acceleration from the beginning up until several minutes into the speech. The degree of acceleration increases throughout the period examined. Quené (2013) notes that there is "only scarce evidence for age-related slowing" in these materials – indeed, what slowing there is takes place before age 58, a trend which then reverses itself from age 58 to age 74.

In summary, the available evidence on age-related changes in speaking tempo, while not particularly extensive or conclusive, suggests that tempo may increase during young adulthood and peak during middle age adulthood. This suggests that, if vowel expansion does take place during young to middle age adulthood, it cannot necessarily be attributed to slowed speaking rate. Our hypothesis that vowel spaces should tend to expand in young to middle age adulthood, in a manner not solely due to vowel lengthening, is broadly consistent with the literature. With this information in mind, we turn to the data for the current study. We test our hypotheses using a corpus of speech samples of eleven talkers age 21-49, recorded once every seven years.

3. Methods

3.1. Data

The formant values and durations analyzed here are those in the "Up" corpus (Gahl et al., 2014), which is based on the "Up" series of documentary films by director Michael Apted. The films feature a set of individuals at seven year intervals over a period of 42 years (Apted, 1977, 1984, 1991, 1998). Most of the speech samples are produced in response to prompts from the director, whose voice can occasionally be heard in the background. The fourteen individuals participated in the filming since 1964 (at age 7), i.e. before they would have been able to give consent, even if procedures for obtaining consent had been in place at the time. Three of the fourteen participants make it clear as adults that they resented being in the films. When the corpus was prepared, Gahl et al. (2014) felt that these individuals should be considered as having withheld or withdrawn their consent and excluded from analysis. The speakers who were included are Andrew, Bruce, John, Lynn, Neil, Nick, Paul, Sue, Suzy, Symon, and Tony. Two of the speakers (John and Symon) refrained from participating in two of the films. The participants for the films were chosen to represent extremes of social class in Britain in 1964, as well as different regional backgrounds: Of the participants included in the corpus, two (Peter and Neil) were born in or near Liverpool, one (Nick) in Yorkshire, and the rest came from London (three attending schools in wealthy neighborhoods, and five attending schools in working-class neighborhoods). One of the five children from working-class backgrounds (Paul) permanently moved to Australia during childhood.

	Andrew	Bruce	John	Lynn	Neil	Nick	Paul	Sue	Suzy	Symon	Tony	Total
21	68	574	606	288	680	358	398	124	204	556	456	4312
28	154	328	0	172	770	786	324	298	242	426	182	3682
35	240	632	726	252	806	470	212	324	474	0	82	4218
42	300	304	0	288	386	470	222	284	374	336	456	3420
49	80	214	450	324	440	486	450	554	328	0	260	3586
Total	842	2052	1782	1324	3082	2570	1606	1584	1622	1318	1436	19218

Table 1: Number of vowel tokens analyzed for each speaker at each age

The corpus consists of audio files and transcripts, time-aligned at the levels of utterance, word, and segment, as well as f0 and vowel formant measurements (in Hz and Bark) of portions of the films, featuring eleven participants (8 male and 3 female) at ages 21 through 49. In our analyses, we used the Bark values; for a discussion of some of the pros and cons of that decision in sociophonetic research, see e.g. Clopper (2009).

The audio files were aligned with an orthographic transcript at the phone level using the Penn Phonetics Lab Forced Aligner Toolkit (Yuan and Liberman, 2008). The phone labels used for the purposes of the temporal alignment were those in the CMU Pronouncing Dictionary (Weide, 1998). Naturally, the individual talkers' pronunciations varied due to their different dialect backgrounds, among other reasons. The phone labels served as labels for speech segments for the purposes of the alignment only. Before the corpus was released, the alignments were hand-checked by a trained phonetician and hand-corrected in cases where the alignment failed, whether due to background noise or other reasons. The IPA labels used in all analyses of vowel formants were based on the (British English) phonetic transcriptions in the CELEX database (Baayen et al., 1995), as follows: An initial mapping from CMU phone labels to CELEX was performed automatically. The transcriptions were then hand-checked by a trained phonetician. For the formant measurements, audio files for each vowel token were extracted, starting 40 ms before the start time and ending 40 ms after the end time of the vowel. The audio was downsampled to 12 KHz and analyzed with the Watanabe and Ueda formant tracker (Ueda et al., 2007). The formant values were those in the analysis frame occurring at the temporal midpoint of the vowel. The

Bark values were calculated using the PhonR package (McCloy, 2016) in R (R Development Core Team, 2008); for discussion of the advantages of using, or refraining from using, Bark values, see Clopper (2009) and Thomas (2013).

The corpus contains 21895 vowel tokens, from 17703 word tokens, representing 2260 word types). We restricted the data set to the lexically stressed tokens of [ɑ, æ, ʋ, ɔ, ɛ, ɪ, i, ʊ, u]. These criteria left 9609 tokens for analysis. Table 1 shows the number of tokens per speaker per age.

	F 1			F 2		
	Correct	Close	Incorrect	Correct	Close	Incorrect
[ɑ]	92.18	0.68	7.14	92.52	2.38	5.10
[æ]	87.28	2.36	10.36	92.00	1.52	6.47
[ɛ]	91.29	1.38	7.32	88.85	1.87	9.28
[i]	87.89	3.50	8.61	64.43	10.89	24.68
[ɪ]	76.29	5.97	17.74	71.44	9.99	18.57
[ɔ]	85.01	2.86	12.13	70.26	18.46	11.27
[u]	80.74	7.05	12.21	72.23	11.31	16.46
[ʊ]	77.58	4.04	18.39	75.78	8.52	15.70
[ʌ]	89.66	1.19	9.15	85.29	3.58	11.13

Table 2: Percentage of automatic formant measurements judged to be correct, approximately correct, and incorrect, based on visual inspection

The automatic formant measurements of the 9609 tokens of stressed vowels were hand-checked by a trained phonetician who was unaware of the research question and hypotheses. We generated image files for each vowel token for this purpose, in which the automatically-extracted F1 and F2 values generated by the formant tracker were marked on a spectrogram. The phonetician inspected these images and categorized each measurement as "correct" (if the formant marker appeared where the phonetician would have placed it), "approximately correct" (if the phonetician felt that the marker was 'slightly' off), or "incorrect" (if the automatic tracking had clearly failed). Instances where the automatic tracking failed mostly fell into three categories: Tokens with significant background noise, cases in which f0 was mistakenly labeled F1 (with F1 being labeled F2), and cases in which what the phonetician judged to be two separate peaks (F1 and F2) being identified as F1, and the label F2 being applied to F3 as a consequence. Table 2 shows the number of correct, approximately correct, and incorrect formant measurements for each formant and vowel. The proportion of F1 measurements judged to be correct ranged from 76% (for [ɪ]) to 92% (for [ɑ]). For F2, the proportion of measurements judged to be correct was similar, ranging from 70% (for [ɔ]) to 92.5% (for [i]). Only those tokens were included in the analysis where both F1 and F2 were judged to be correct, leaving a total of 6642 vowel tokens for analysis.

3.2. Statistical treatment of the data

The analysis presented here is the result of a process of familiarization with the data, in which initial survey models were gradually updated as we delved deeper into the literature on aging and on the factors determining vowel spaces. We fitted two models, one without age as a predictor, which serves as our baseline model, and one model that does include age. Comparison of the two models allows us to establish whether the vowel space indeed increases with age.

Treatment coding was used for all factors, meaning that the estimates for each factor level indicate comparisons to a reference level dummy-coded as 0. Vowel was coded as a factor with [ɑ] as the reference level; the reference level for Formant was F1, meaning that, for example, the estimated coefficient for 'Vowel æ' specifies the magnitude of the difference in F1 between [æ] and [ɑ] differed in F1. Duration was log-transformed, to approach normality more closely and as is common in research on duration and other continuous variables (e.g. lexical frequency) in which a given absolute difference (e.g. 20 milliseconds or "50") must be interpreted differently in lower vs. higher ranges of the scale in question. Age, a numeric variable, was centered, by subtracting the overall mean age from each observation, and scaled, by subtracting the centered observations by their standard deviation.

Our initial models were linear mixed effects regression models, fitted using the package lme4 (Bates et al., 2015) in R (R Development Core Team, 2008). Model criticism revealed that these initial models failed to meet the assumptions underlying linear mixed effects regression (see Supplementary Material for details). We therefore used the framework of the generalized additive model (GAM, Wood, 2017) to fit a mixed-effects model to the data, using the `gam` function from the `mgcv` package (Wood, 2011). For a tutorial on using GAMs in phonetic research, see Wieling (2018). Recent publications using GAMs in research related to the present paper include Tomaschek et al. (2018b).

The fixed effect part of the linear predictor for formant frequency,

```
(Sex + Duration) * Formant * Vowel,
```

included all main effects, pairwise interactions, and three-way interactions of these four predictors with the exception of two- and three-way interactions involving `sex` and `duration`, which did not receive support in any of the models we considered. All other interactions were included to capture vowel-specific effects of duration and sex-related F1-F2 differences (Diehl et al., 1996; Munson and Babel, 2016; Weirich and Simpson, 2014; Fant, 1966; Hillenbrand et al., 1995; Johnson, 2006; Byrd, 1994).

Model summaries and information on the treatment of the data and statistical models can be found in Appendix A. Here, we primarily focus the description of the findings on the patterns reflecting overall properties of the talkers' F1/F2 space and on the effects of age and vowel duration on vowel position in that space.

As random-effect factors we included `speaker`, the phone preceding the vowel (`phone_pre`) as well as the phone following the vowel (`phone_post`). The preceding and following phones are included to take into account the consequences of co-articulation with preceding and following consonants (Hillenbrand et al., 1995; van Santen, 1992; Stevens and House, 1963). Interactions of `formant` by `speaker`, `phone_pre` and `phone_post` were included to allow effects to be expressed not only on F1 but also on F2. Exploratory analyses furthermore revealed strong evidence for by-speaker random slopes for `duration`.

Details of pronunciation may be word-specific (see e.g. Pierrehumbert, 2002; Gahl, 2008; Hay et al., 2015). Moreover, lexical choice may systematically be related to age (Kavé and Nussbaum, 2012; Horton et al., 2010). One might therefore wish to control for lexical effects on vowel characteristics, as

F1 and F2 of a target vowel might depend on the word in which the vowel appeared. Although we considered including `word` as a random effect factor, a model including both `word` and the preceding and following consonants would be overspecified: For more than half of the word types, our database provides only 1 datapoint. A model with `word`, `phone_pre` and `phone_post` invests three coefficients for each word type, leading to overdetermination for 34.85% of the words. However, word-specific effects will be largely covered by the random effects for `phone_pre` and `phone_post`, as the number of unique combinations of `phone_pre` and `phone_post` (1295) accounts for 82.96% of the number of word types (1561).

4. Results

4.1. Models of formant frequencies

The model summaries for the full dataset appear in Tables S1 and S2 in the Supplementary Material. Here, we first focus on the hand-curated set, shown in Tables A1 (parametric terms) and A2 (smooth terms) in Appendix A.

The estimates of Vowel and its interactions with Formant capture the position of vowels in the F1/F2 plane: As one might expect, the vowel with the lowest estimated F1 is [i] (-4.84, $p < .0001$). Vowel position along the F2 axis is reflected in the effect of Formant and in the estimates of the interaction between Vowel and Formant, which is again largest in the case of the front vowel [i] (11.26, $p < .0001$). Comparison of the GAMMs with and without `Age` showed that the model with Age achieved an increase in model fit. The estimates for the effect of Age on the first formant are significant for the vowels [i, ɪ, ɔ, ʊ], and [u]: With increasing talker age, there was a significant decrease in F1 (i.e. vowel raising) relative to the reference level [ɑ]. This decrease was strongest for the high vowels [i], [ɪ], and [ʊ]). The effect of Age on the second formant is reflected in the three-way interaction of Age with Formant and Vowel. The model estimates are significant for the vowels [i],[ɪ, and [u]. In all three of these vowels, F2 increases with increasing talker age.

The pattern of results based on the full set of automatic measurements was very nearly the same as the results based on the hand-curated set, with very few exceptions that formed a very clear pattern: Among the 72 parametric estimates, there were four that reached significance in the full set, but not in the subset, and two that reached significance only in the subset. The vowels implicated in these six exceptions were [ɑ, u, ʊ, ʊ], and [ɔ], all of which are characterized by small differences between F1 and F2 (c.f. e.g. Hillenbrand et al. 1995 and Peterson and Barney 1952). As pointed out in Watson and Harrington (1999), vowels in which F1 and F2 are close together make challenging targets for automatic formant tracking.

4.2. The relationship between vowel duration, talker age, and F1/F2

Interestingly, the effects of duration and talker age were similar in some ways: The vowels [i, ɪ, ɔ, ʊ], and [u] were raised with increasing duration, as with increasing age. Similarly, F2 increased with increasing duration and age for the vowels [i, ɪ, ɛ, æ], and [u]. To explore the relationship between vowel duration and talker age further, we first asked whether the relationship between Duration and Age was similar along the whole range of vowel duration and talker age. It could be the case, for example, that the expansion of vowel spaces was only observable in vowels above a certain minimum duration, if articulatory speed in very short vowels was simply not high enough to enable talkers, regardless of age, to reach highly peripheral places of articulation. Alternatively, the hypothesized shift of vowel productions towards the periphery of the space might only be observable in short vowels, if younger talkers' productions were as peripheral as those of their older selves' given more time. A third possibility might be for the difference between "young" and "old" vowel spaces to be constant across vowel durations.

To answer these questions, we examined the change in vowel space expansion using convex hulls. A convex hull of a set of points in a Euclidean plane (the F1/F2 plane in our case) is the smallest convex set (or 'region') that contains the points. Informally speaking, a convex hull might be pictured as the area enclosed by a rubber band wrapped around pegs on a peg board. Figure 3 shows the model predictions (based on the subset of measurements that were found to be correct) for vowel spaces at five evenly spaced values of vowel durations, from the minimum duration (0%) to the maximum (100%). The predicted vowel spaces are depicted as convex hulls fitted around the fixed-effects estimates, i.e. the predictions purely based on the identity of the vowel, formant (F1 or F2), talker age, and the duration of the token, setting aside the by-talker and by-context adjustments. In each panel, we show two sets of predictions, one for age 21 (dark grey), and one for age 49 (shown in light grey). The estimates for ages 28, 35, and 42 are not shown, but fall between the estimates for ages 21 and 49.

For the research questions of our study, Figure 3 can be regarded as the crucial summary of the model of formant values. The effect of vowel duration manifests itself as a considerable increase in the overall space from the short-

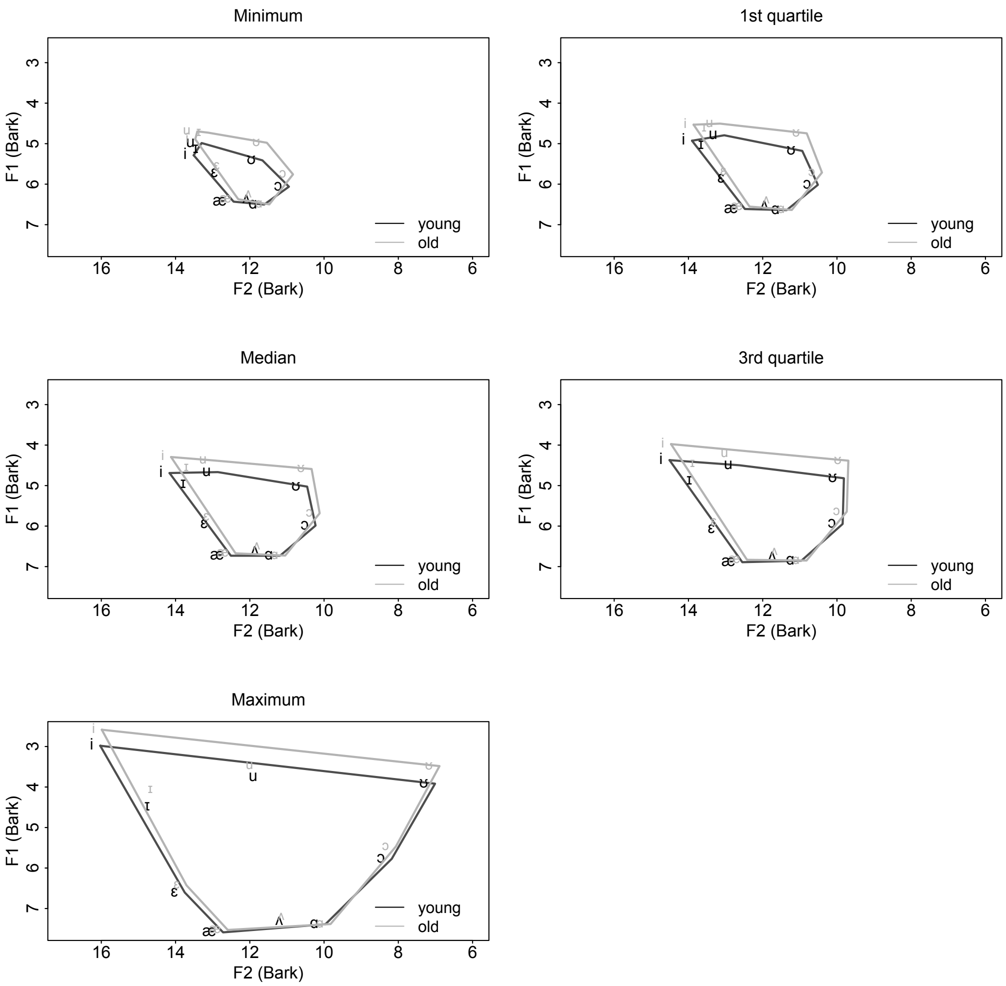

Figure 3: Model predictions for vowel spaces at five evenly spaced values of vowel durations, from the minimum duration (0%) to the maximum (100%). Dark grey indicates the predictions at age 21, i.e. the youngest age represented in the corpus. Light grey indicates the predictions for age 49, the oldest age represented in the corpus.

est durations (leftmost panel) to the longest durations (rightmost panel). Three patterns are worth noting here: First, the area increases as duration increases - consistent with the general observation that longer vowels tend to occupy larger F1/F2 spaces. Second, within each panel, the area is larger at age 49 than at age 21: Age is predictive of vowel space expansion when controlling for vowel duration. The pattern of vowel spaces being larger in the older speakers holds across the entire span of vowel duration: In each panel, the predicted space is larger for age 49 than for age 21. Third, the relative differences between age 21 and 49 are proportionally largest for shorter durations, as opposed to remaining constant across the range of durations.

5. Discussion

We hypothesized that, as talkers move through young-to-middle age adulthood, they speed up articulatory movements as necessary to maintain articulatory targets even of short vowels. Our modeling results are consistent with that hypothesis. We found that F1 and F2 of stressed vowels changed across the five speech samples recorded at ages 21 through age 49, in a manner suggesting that vowel space size increased as talkers got older. There was a complex relationship between vowel spaces, vowel duration, and age, such that the relative differences between age 21 and 49 were proportionally largest for short vowel durations. We observed essentially the same pattern of results in the raw output of an automatic formant tracker and in the subset of formant values judged to be correct based on visual inspection by a research assistant.

The fact that the difference between the vowel spaces at age 21 vs. 49

depended on vowel duration is also consistent with our hypothesis. At the same time, the fact that the effect of age persists even in the longest vowel tokens suggests that the peripheral realizations are not solely a matter of an increase in the sheer ability to reach extreme places of articulation quickly enough: If that were the case, then the younger talkers should reach maximally peripheral positions given enough time. As a result, the difference between young and old talkers should disappear in long vowels in that scenario. The presence of the effect of age in even the longest vowels suggests that target realization is not simply a matter of how much time is available in a specific utterance.

5.1. Limitations

The current study has a number of limitations, due to the small size of the corpus, and due to the nature of the recordings. Some of these limitations are methodological in nature. An anonymous reviewer points out that comparing our findings to previous studies, and comparing previous studies to one another, is problematic, due to methodological differences. This problem is highlighted in Watson and Evans (2016), who found that the same diachronic corpora, analyzed using different forced-alignment methods, led to different conclusions about sound change. A related limitation results from our decision to analyze formant patterns at a single point (at or near the temporal midpoint) in each vowel token. Follow-up analyses of spectral *trajectories* would be highly desirable, given that our interpretation of the results relates not just to (acoustic or articulatory) target positions, but to velocity and trajectory of lingual movement.

The size of our data set prevents us from estimating word-specific, as opposed to vowel-specific, effects of age or reaching firm conclusions about the nature of the articulatory (or perceptual) targets. One interpretation of the changes we observe is that (articulatory or perceptual) targets for vowels come to be more peripheral due to a given vowel having been spoken more often as speakers age. That is not the only possible interpretation of our results. In fact, we expect the targets of learning to be words, rather than vowels. That expectation is consistent with a body of research on "word-specific phonetics" (Pierrehumbert, 2002; Johnson, 1997, 2007), including research tracking word-specific shifts in pronunciation over time (Hay et al., 2015).

In analyses of possible effects of talker age, taking into account such lexically-specific patterns may be particularly appropriate: Adult vocabulary size increases over the life time (Keuleers et al., 2015), and lexical choices differ across age groups, such that talkers tend to use higher proportions of low-frequency words with increasing age (Kavé and Nussbaum, 2012; Horton et al., 2010; Meylan and Gahl, 2014; Moscoso del Prado Martín, 2016). Elsewhere, we have argued that such changes in lexical choice contribute to the apparent slow-down in lexical retrieval associated with age (Ramscar et al., 2013). Lexical frequency in turn affects word durations in young and old adults (Moers et al., 2016). Therefore, one would expect longer word durations, and perhaps vowel durations, in older vs. younger adults, as an indirect consequence of increased vocabulary size and use of low-frequency words: Changes in word choice could conceivably lead to changes in vowel spaces.

In our analyses, we controlled for phonological context (the segments preceding and following each target); as mentioned in section 3.2 above, the random effects encoding the consonants flanking the target vowels effectively control for word choice. As a simple follow-up test of whether there was a systematic relationship between word frequency and age in our corpus, we compared the average word frequency at each age to the overall average word frequency. There was no consistent pattern: Words used at age 21 were more frequent, compared to the overall average frequency; but words used at age 49 were not less frequent, compared to the overall average; ages 28, 35, and 42 also did not indicate a clear trend. It seems unlikely, therefore, that the observed change in vowel spaces results from changes in word choice. We nevertheless believe that the respective roles of lexical and phonological context in vowel realization deserve further investigation.

5.2. Relationship to previous literature on aging and speech

The role of vowel duration in the observed patterns may shed light on the relationship between our findings and previous literature, and on the interpretation of previous findings. If vowel spaces expand during young to middle age adulthood and contract in elderly talkers, the question arises when age-related vowel centralization might begin, and whether changes during middle age adulthood preshadow changes that take place later in life. We believe that the evidence about age-related changes in vowel spaces does not at present answer that question, due to several methodological problems.

The first such problem is that, as noted earlier, few studies control for vowel duration. Our finding that age effects were similar at all levels of duration underscores the need to control for this variable: For example, comparing formant values for long (and hence, tending to be peripheral, all else being equal) vowels produced by young talkers vs. short (and hence, central) vowels produced by old talkers would make vowels appear to centralize with talker age. Conversely, comparing long vowels produced by old talkers with short vowels produced by young talkers would vastly exaggerate the effect we observed, of vowels coming to be more peripheral with talker age. Complicating matters further is the fact that age and speaking rate are not independent, partly due to the relationship between talker age and word choice. Thus, failing to control for vowel duration renders comparisons across studies problematic in several ways.

A related methodological problem in comparisons across studies without controling for vowel duration arises due to differences between vowels in running speech vs. sustained vowels produced in isolation. A third problem is that, as Watson and Evans (2016) point out, comparisons across studies using different automatic formant extraction cannot be taken at face value. We also note that many patterns reported for talkers over the age of approximately 60 or 70 may reflect the length of time in retirement and the attendant reduction in social networks (Dufouil et al., 2014; Ramscar et al., 2014). As far as we can tell, none of the studies we surveyed controlled for this variable.

Finally, differences in pronunciation across age groups also reflect sociolinguistic factors such as group affiliation, persona, and style (Docherty, 2007; Docherty and Mendoza-Denton, 2012; Wagner, 2012; Sankoff, 2013; Eckert, 2004; Rickford and Price, 2013). Our observations could be similarly interpreted as reflecting changes in 'personal and social expectancies', along with geographical moves, the use of regional dialects, and social status (Buchstaller et al., 2017; Clopper et al., 2017). Consistent with this interpretation, the individuals featured in the films on which our samples are based change in many ways besides speech as their outlook on life changes. To name just one relevant pattern, there are changes in posture and apparent mood. However, we think it unlikely that the observed change in vowel space size over time is entirely due to factors other than age: Each of the talkers underwent different changes in mood and geographic location, as well as in education level, income, and many other factors. We cannot rule out the possibility that individual differences in additional highly heterogeneous variables conspired to produce a spurious effect of age in the group as a whole, but we believe the probability of such a conspiracy to be small.

5.3. Methodological implications

The similarities between the models of the full set of automatic formant measurements vs. the subset that were found to be correct should be of interest to researchers needing to make decisions about how to allocate resources for segmentation and alignment at the phone level, hand-checking, and hand-correcting formant measurements. As noted earlier, certain vowels - here, those in which F1 and F2 are close together - are particularly vulnerable to tracking errors, consistent with what has been found in systematic investigations of formant analyses (Watson and Harrington, 1999). Moreover, the ability of statistical models of formant measurements to pick up on patterns in data, and the risk of flagging spurious patterns as "significant", reflect factors other than the error rate in formant tracking: The distribution of (correct and errorful) measurements, along with the error rate, determine which errors are harmful (see Lee 1987 for discussion). For many specific research projects, uncorrected measurements may yield results that are sufficiently accurate for answering the question of interest.

5.4. Implications for the concept of 'phonetic reduction'

Recent years have seen a great deal of interest in details of pronunciation as a tool in psycholinguistic modeling and hypothesis testing. One assumption in much of that literature has been that durational shortening, vowel formants, and articulatory range can all safely be regarded as manifestations of a single, broad pattern of across-the-board phonetic reduction. An immediate implication of the present findings is that vowel shortening cannot be taken as *prima facie* evidence for reduction generally. Conversely, vowel peripheralization does not imply lengthening or 'strengthening' generally.

The idea that temporal reduction should generally lead to articulatory reduction rests on the assumption that the speed of articulatory movements stays constant. Often, that is not the case. Research on vocal aging underscores the need for a nuanced view of durational and articulatory aspects of pronunciation, and of the role of age and learning in both. Our findings highlight the partial independence of vowel duration and vowel formants, and hence the risk inherent in regarding both as instantiations of across-the-board reduction or enhancement.

6. Conclusion

Inspired in part by recent research on the effects of learning on articulatory movement (Tomaschek et al., 2018a), we hypothesized that vowel formants should change across young to middle-age adulthood producing an overall increase in vowel space size. An analysis using Generalized Additive

Models (GAMs, Wood 2017) bore this out. The increase in vowel space size was not fully attributable to an increase in vowel duration. That latter finding is problematic for the assumption, often made in research on phonetic reduction and strengthening, that duration and position in vowel space covary.

At a more general level, our hypotheses were inspired by recent work showing that age-related change in language processing need not reflect cognitive decline, but, to the contrary, reflect increasing expertise (Ramscar et al., 2013). Our findings are consistent with recent findings on the accumulation of knowledge over the lifespan and its consequences for the cognition of healthy aging.

Acknowledgments

We are grateful to Ronald Sprouse and Sofea Dil for research assistance. This research was supported by an Alexander von Humboldt Professorship (grant number 1141527) and ERC advanced grant (number 742545) awarded to the second author.

Appendix A: Regression information

As positions in the vowel space are determined by two axes, one for F1 and one for F2, we opted for a model predicting formant frequency (on the bark scale) while including a factorial contrast distinguishing between F1 (the reference level) and F2, using treatment coding (see e.g. Faraway, 2006, for this way of modeling paired response variables). The reference level for the factor Vowel was ɑ.

Our initial models were linear mixed effects regression models, fitted using the package lme4 (Bates et al. 2015) in R. We fitted the baseline model with maximum likelihood to enable subsequent comparison with the model including `Age` as predictor. Model criticism revealed that the residuals of this baseline model did not follow a Gaussian distribution, but rather a scaled t-distribution. We therefore made use of the framework of the generalized additive model (GAMs, Wood 2017) to fit a mixed-effects model to the data, using the `gam` function from the `mgcv` package (Wood, 2011), which makes it possible to properly model the scaled residuals as following a t-distribution. That is,

$$\frac{y - \mu}{\sigma} \sim t_{\nu},$$

where μ is determined by the linear predictor while σ and ν are parameters that are estimated along with the other parameters of the model (Wood, 2016).

The model that we fitted to the data is very similar to an LMER, the main difference being that random effects are estimated by imposing a ridge penalty.

For comparing models with different fixed effects, maximum likelihood was used; but for the final model (not used for model comparison) we refit with restricted maximum likelihood: For assessing `Age` as predictor of vowel formants, we fitted a second GAMM, again using maximum likelihood, including a simple effect of `Age`, by-speaker random slopes for `Age`, and the interaction of `Age` with the term representing the interaction of vowel with formant. We then refitted the models with and without `Age` with restricted maximum likelihood. In order to allow the effect of `Age` to vary by `speaker`, we also included by-speaker random slopes for `Age`.

Comparison of the GAMMs with and without `Age` showed that the model with Age achieved an increase in model fit as assessed by a difference in ML scores equal to 39.494, achieved at the cost of 18 degrees of freedom. As the difference follows a chi-squared distribution with 18 degrees of freedom, the increase in model complexity is outweighed by the increase in model fit ($p \ll 0.0001$). We then refitted the model that included `Age` with restricted maximum likelihood. A quantile-quantile plot of the residuals of this model revealed only minor deviations from normality, indicating that the scaled t-distribution (with estimated parameters 5.266 and 0.579) was appropriate.

Tables A 3 and A 4 show the parametric and smooth terms of the model, respectively. Estimates that are mentioned in the description of the results are in **boldface**. Only some salient coefficients are discussed in the text: We have refrained from discussing how higher-order interactions further modulate the differences highlighted by the salient coefficients.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	7.6541	0.2317	33.0341	0.0000
Sex *male*	**-0.6629**	**0.1597**	**-4.1500**	**0.0000**
Formant *F2*	1.8947	0.3276	5.7834	0.0000
Vowel æ	0.1904	0.2249	0.8466	0.3972
ɛ	-0.8584	0.2461	-3.4881	0.0005
i	**-4.8432**	**0.2362**	**-20.5077**	**0.0000**
ɪ	-3.2098	0.2380	-13.4885	0.0000
ɔ	-1.8433	0.2208	-8.3498	0.0000
u	-3.9278	0.2414	-16.2740	0.0000
ʊ	-3.8259	0.4592	-8.3324	0.0000
ʌ	**-0.1285**	**0.2384**	**-0.5388**	**0.5900**
Duration	0.2812	0.0860	3.2692	0.0011
Age	-0.0179	0.0456	-0.3928	0.6944
Sex *male*:Formant *F2*	**-0.0230**	**0.2258**	**-0.1021**	**0.9187**
Sex *male*:Vowel æ	-0.0347	0.1143	-0.3035	0.7615
Sex *male*: ɛ	0.1272	0.1168	1.0891	0.2761
Sex *male*: i	-0.1014	0.1177	-0.8619	0.3887
Sex *male*: ɪ	0.1003	0.1154	0.8690	0.3849
Sex *male*: ɔ	-0.1240	0.1160	-1.0687	0.2852
Sex *male*: u	-0.0477	0.1274	-0.3739	0.7085
Sex *male*: ʊ	-0.0013	0.1870	-0.0071	0.9944
Sex *male*: ʌ	-0.1954	0.1182	-1.6538	0.0982

Continued on next page

Table 3 – continued from previous page

	Estimate	Std. Error	z value	Pr(>|z|)
Formant *F2*:Vowel æ	2.7112	0.3178	8.5308	0.0000
Formant *F2*: ɛ	4.8811	0.3479	14.0311	0.0000
Formant *F2*: i	**11.2556**	**0.3339**	**33.7096**	**0.0000**
Formant *F2*: ɪ	7.9906	0.3362	23.7659	0.0000
Formant *F2*: ɔ	-0.0269	0.3121	-0.0862	0.9313
Formant *F2*: u	5.6919	0.3411	16.6848	0.0000
Formant *F2*: ʊ	0.6534	0.6491	1.0066	0.3141
Formant *F2*: ʌ	1.1783	0.3371	3.4956	0.0005
Formant *F2*:Duration	-0.7835	0.1216	-6.4416	0.0000
Vowel æ:Duration	0.0768	0.0947	0.8111	0.4173
ɛ:Duration	-0.0085	0.0994	-0.0855	0.9319
i:Duration	**-0.9873**	**0.0988**	**-9.9981**	**0.0000**
ɪ:Duration	**-0.4793**	**0.0969**	**-4.9459**	**0.0000**
ɔ:Duration	**-0.3655**	**0.0923**	**-3.9619**	**0.0001**
u:Duration	**-0.6538**	**0.1000**	**-6.5405**	**0.0000**
ʊ:Duration	**-0.7302**	**0.1616**	**-4.5181**	**0.0000**
ʌ:Duration	0.0103	0.0990	0.1038	0.9173
Formant *F2*:Age	-0.0501	0.0615	-0.8146	0.4153
Vowel æ:Age	-0.0204	0.0481	-0.4239	0.6717
ɛ:Age	-0.0632	0.0484	-1.3060	0.1916
i:Age	**-0.1393**	**0.0502**	**-2.7772**	**0.0055**
ɪ:Age	**-0.1459**	**0.0486**	**-3.0026**	**0.0027**

Continued on next page

Table 3 – continued from previous page

	Estimate	Std. Error	z value	Pr(>|z|)
ɔ:Age	**-0.1071**	**0.0482**	**-2.2243**	**0.0261**
u:Age	**-0.1027**	**0.0523**	**-1.9652**	**0.0494**
ʊ:Age	**-0.1558**	**0.0712**	**-2.1878**	**0.0287**
ʌ:Age	-0.0400	0.0497	-0.8062	0.4201
Sex *male*:Formant *F2*:Vowel æ	-0.1066	0.1615	-0.6600	0.5092
Sex *male*:Formant *F2*: ɛ	-0.1719	0.1651	-1.0414	0.2977
Sex *male*:Formant *F2*: i	-0.0375	0.1663	-0.2256	0.8215
Sex *male*:Formant *F2*: ɪ	-0.2033	0.1631	-1.2466	0.2125
Sex *male*:Formant *F2*: ɔ	-0.0803	0.1640	-0.4898	0.6243
Sex *male*:Formant *F2*: u	**-0.3891**	**0.1801**	**-2.1604**	**0.0307**
Sex *male*:Formant *F2*: ʊ	-0.0149	0.2644	-0.0563	0.9551
Sex *male*:Formant *F2*: ʌ	-0.0460	0.1671	-0.2755	0.7830
Formant *F2*:Vowel æ:Duration	**0.5146**	**0.1338**	**3.8457**	**0.0001**
Formant *F2*: ɛ:Duration	**0.8413**	**0.1406**	**5.9855**	**0.0000**
Formant *F2*:i:Duration	**2.2627**	**0.1396**	**16.2047**	**0.0000**
Formant *F2*: ɪ:Duration	**1.3826**	**0.1370**	**10.0954**	**0.0000**
Formant *F2*: ɔ:Duration	0.0181	0.1305	0.1389	0.8896
Formant *F2*: u:Duration	**0.6416**	**0.1413**	**4.5396**	**0.0000**
Formant *F2*: ʊ:Duration	-0.1924	0.2285	-0.8422	0.3997
Formant *F2*: ʌ:Duration	0.2197	0.1399	1.5701	0.1164
Formant *F2*:Vowel æ:Age	0.0249	0.0680	0.3657	0.7146
Formant *F2*: ɛ:Age	0.0965	0.0684	1.4112	0.1582

Continued on next page

Table 3 – continued from previous page

	Estimate	Std. Error	z value	Pr(>|z|)
Formant *F2*: i:Age	**0.1745**	**0.0709**	**2.4605**	**0.0139**
Formant *F2*: ɪ:Age	**0.1704**	**0.0687**	**2.4814**	**0.0131**
Formant *F2*: ɔ:Age	0.1183	0.0681	1.7383	0.0822
Formant *F2*: u:Age	**0.1924**	**0.0738**	**2.6060**	**0.0092**
Formant *F2*: ʊ:Age	0.1616	0.1006	1.6063	0.1082
Formant *F2*: ʌ:Age	0.0759	0.0702	1.0815	0.2795

Table 3: Summary of the GAM model of first and second formants, hand-checked measurements, parametric coefficients

Supplementary material: Results based on the full set of automatic formant measurements

Comparison of the GAMMs with and without `Age` showed that the model with Age achieved an increase in model fit, assessed by a difference in ML scores equal to 46.928, at the cost of 18 degrees of freedom. As the difference follows a chi-squared distribution with 18 degrees of freedom, the increase in model complexity is outweighed by the increase in model fit ($p \ll 0.0001$). We then refitted the model that included `Age` with restricted maximum likelihood. A quantile-quantile plot of the residuals of this model revealed only minor deviations from normality, indicating that the scaled t-distribution (with estimated parameters 4.827 and 0.709) was appropriate.

Comparison of the GAMMs based on the full set of automatic measurements vs. the subset found to be correct revealed very few differences, as follows: Three effects that were significant in the complete dataset were no longer significant after removal of the problematic measurements: (1) The interaction of Sex with Vowel for the vowel [ɔ] as produced by the male talkers, which suggested that F1 was lower in that vowel, compared to the reference vowel [ɑ]; no such interaction was present in the subset; (2) The interaction of Formant with Vowel, indicating that F2 was higher in [ʊ] compared to [ɑ]; and (3) The interaction of Formant with Vowel suggesting that F2 in [ʌ] was higher (1.28), compared to [ɑ]. Two effects that were non-significant in the complete dataset rose to significance in the hand-curated set: (1) the simple effect of Duration (indicating that longer duration was associated with higher F1 in the reference vowel [ɑ], i.e. a more peripheral position in F1/F2 space); (2) the three-way interaction of Sex and Formant with Vowel, indicating that, for male talkers, F2 in [u] was lower (-0.39, $p < .03$) compared to the reference level.

References

Michael Apted. *21 Up.* ITV Granada Television, 1977.

Michael Apted. *28 Up.* ITV Granada Television, 1984.

Michael Apted. *35 Up.* ITV Granada Television, 1991.

Michael Apted. *42 Up.* BBC, 1998.

Matthew Aylett and Alice Turk. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The*

Journal of the Acoustical Society of America, 119(5):3048–3058, 2006. doi: 10.1121/1.2188331.

R Harald Baayen, Richard Piepenbrock, and Leon Gulikers. The celex lexical database (release 2). *Distributed by the Linguistic Data Consortium, University of Pennsylvania*, 1995.

Ronald J Baken. The aged voice: a new hypothesis. *Journal of Voice*, 19(3): 317–325, 2005. doi: 10.1016/j.jvoice.2004.07.005.

Ronald J Baken and Robert F Orlikoff. *Clinical measurement of speech and voice*. Cengage Learning, 3 edition, 2000.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi: 10.18637/jss.v067.i01.

Alan Bell, Jason M Brenier, Michelle Gregory, Cynthia Girand, and Dan Jurafsky. Predictability effects on durations of content and function words in conversational english. *Journal of Memory and Language*, 60(1):92–111, 2009. doi: 10.1016/j.jml.2008.06.003.

Barbaranne J Benjamin. Phonological performance in gerontological speech. *Journal of Psycholinguistic Research*, 11(2):159–167, 1982.

Stephen D Bier, Catherine I Watson, and Clare M McCann. Dynamic measures of voice stability in young and old adults. *Logopedics Phoniatrics Vocology*, 42(2):51–61, 2017.

Ann R Bradlow. Confluent talker-and listener-oriented forces in clear speech production. *Laboratory phonology*, 7:241–273, 2002. doi: 10.1515/9783110197105.241.

Angelika Braun and Stefan Friebis. Phonetic cues to speaker age: A longitudinal study. *Formal Linguistics and Law*, 212:141–162, 2009.

James Brink, Richard Wright, and David B Pisoni. Eliciting speech reduction in the laboratory: Assessment of a new experimental method. *Research on Spoken Language Processing Progress Report No*, 22, 1998.

Isabelle Buchstaller, Anne Krause, Anja Auer, and Stefanie Otte. Levelling across the life-span?: Tracing the face vowel in panel data from the north east of england. *Journal of Sociolinguistics*, 21(1):3–33, 2017. doi: 10.1111/josl.12227.

Dani Byrd. Relations of sex and dialect to reduction. *Speech Communication*, 15(1-2):39–54, 1994. doi: 10.1016/0167-6393(94)90039-6.

Cynthia G Clopper. Computational methods for normalizing acoustic vowel data for talker differences. *Language and Linguistics Compass*, 3(6):1430–1442, 2009.

Cynthia G Clopper, Jane F Mitsch, and Terrin N Tamati. Effects of phonetic reduction and regional dialect on vowel production. *Journal of Phonetics*, 60:38–59, 2017.

Violet O Cox and Mark Selent. Acoustic and respiratory measures as a function of age in the male voice. *Journal of Phonetics & Audiology*, 2015, 2015. doi: 10.4172/jpay.100010.

Wivine Decoster and Frans Debruyne. Longitudinal voice changes: facts and interpretation. *Journal of Voice*, 14(2):184–193, 2000. doi: 10.1016/S0892-1997(00)80026-0.

Randy L Diehl, Björn Lindblom, Kathryn A Hoemeke, and Richard P Fahey. On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of phonetics*, 24(2):187–208, 1996. doi: 10.1006/jpho.1996.0011.

Gerard Docherty and Norma Mendoza-Denton. Speaker-related variation–sociophonetic factors. *The Oxford Handbook of Laboratory Phonology*, pages 44–60, 2012.

Gerard J Docherty. Speech in its natural habitat: accounting for social factors in phonetic variability. *Laboratory phonology*, 9:1–36, 2007.

Christopher Dromey and Sarah Scott. The effects of noise on speech movements in young, middle-aged, and older adults. *Speech, Language and Hearing*, pages 1–9, 2016. doi: 10.1080/2050571X.2015.1133757.

C. Dufouil, E. Pereira, and G. et al. Chêne. Older age at retirement is associated with decreased risk of dementia. *European Journal of Epidemiology*, 29:353–361, 2014.

Penelope Eckert. The meaning of style. In *Salsa 11. Texas Linguistics Forum. 47.*, 2004.

Julie Traub Eichhorn, Raymond D Kent, Diane Austin, and Houri K Vorperian. Effects of aging on vocal fundamental frequency and vowel formants in men and women. *Journal of Voice*, 32(5):644.e1–644.e9, 2018. ISSN 0892-1997. doi: 10.1016/j.jvoice.2017.08.003.

Gunnar Fant. A note on vocal tract size factors and non-uniform f-pattern scalings. *Speech Transmission Laboratory Quarterly Progress and Status Report*, 1:22–30, 1966.

Julian J. Faraway. *Extending Linear Models with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models.* Chapman & Hall/CRC, Boca Raton, FL, 2006.

Angela Fink and Matthew Goldrick. The influence of word retrieval and planning on phonetic variation: Implications for exemplar models. *Linguistics Vanguard*, 1(1):215–225, 2015.

Annalise R Fletcher, Megan J McAuliffe, Kaitlin L Lansford, and Julie M Liss. The relationship between speech segment duration and vowel centralization in a group of older speakers. *The Journal of the Acoustical Society of America*, 138(4):2132–2139, 2015. doi: 10.1121/1.4930563.

Marios Fourakis. Tempo, stress, and vowel reduction in American English. *The journal of the Acoustical society of America*, 90(4):1816–1827, 1991. doi: 10.1121/1.401662.

Robert Allen Fox and Ewa Jacewicz. Dialect and generational differences in vowel space areas. In *Third ISCA Workshop on Experimental Linguistics*, 2010.

Susanne Gahl. Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84(3):474–496, 2008. doi: 10.1353/lan.0.0035.

Susanne Gahl and Susan M Garnsey. Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, pages 748–775, 2004. doi: 10.1353/lan.2004.0185.

Susanne Gahl and Julia F Strand. Many neighborhoods: Phonological and perceptual neighborhood density in lexical production and perception. *Journal of Memory and Language*, 2016. doi: 10.1016/j.jml.2015.12.006.

Susanne Gahl, Yao Yao, and Keith Johnson. Why reduce? phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66(4):789–806, 2012. doi: 10.1016/j.jml.2011.11.006.

Susanne Gahl, Emily Cibelli, Kathleen Hall, and Ronald Sprouse. The 'Up' corpus: A corpus of speech samples across adulthood. *Corpus Linguistics and Linguistic Theory*, 10(2):315–328, 2014. doi: 10.1515/cllt-2013-0023.

Thomas Gay. Effect of speaking rate on vowel formant movements. *The journal of the Acoustical society of America*, 63(1):223–230, 1978. doi: 10.1121/1.381717.

Elizabeth Gordon, Lyle Campbell, Jennifer Hay, Margaret Maclagan, Andrea Sudbury, and Peter Trudgill. *New Zealand English: its origins and evolution*. Cambridge University Press, 2004. doi: 10.1017/cbo9780511486678.

Huiwen Goy, David N Fernandes, M Kathleen Pichora-Fuller, and Pascal van

Lieshout. Normative voice data for younger and older adults. *Journal of Voice*, 27(5):545–555, 2013. doi: 10.1016/j.jvoice.2013.03.002.

James D Harnsberger, Richard Wright, and David B Pisoni. A new method for eliciting three speaking styles in the laboratory. *Speech communication*, 50(4):323–336, 2008. doi: 10.1016/j.specom.2007.11.001.

Jonathan Harrington. An acoustic analysis of âĂŸhappy-tensingâĂŹin the queen's christmas broadcasts. *Journal of Phonetics*, 34(4):439–457, 2006.

Jonathan Harrington, Sallyanne Palethorpe, Catherine I Watson, et al. Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers. In *Interspeech*, pages 2753–2756, 2007.

Jennifer B Hay, Janet B Pierrehumbert, Abby J Walker, and Patrick LaShell. Tracking word frequency effects through 130years of sound change. *Cognition*, 139:83–91, 2015. doi: 10.1016/j.cognition.2015.02.012.

James Hillenbrand, Laura A Getty, Michael J Clark, and Kimberlee Wheeler. Acoustic characteristics of american english vowels. *The Journal of the Acoustical society of America*, 97(5):3099–3111, 1995. doi: 10.1121/1.411872.

Harry Hollien and Thomas Shipp. Speaking fundamental frequency and chronologic age in males. *Journal of Speech, Language, and Hearing Research*, 15(1):155–159, 1972. doi: 10.1044/jshr.1501.155.

Harry Hollien, Patricia A Hollien, and Gea de Jong. Effects of three parameters on speaking fundamental frequency. *The Journal of the Acoustical Society of America*, 102(5):2984–2992, 1997. doi: 10.1121/1.420353.

William S Horton, Daniel H Spieler, and Elizabeth Shriberg. A corpus analysis of patterns of age-related change in conversational speech. *Psychology and aging*, 25(3):708, 2010. doi: 10.1037/a0019424.

Ewa Jacewicz, Robert Allen Fox, and Lai Wei. Between-speaker and within-speaker variation in speech tempo of american english. *The Journal of the Acoustical Society of America*, 128(2):839–850, 2010. doi: 10.1121/1.3459842.

Ewa Jacewicz, Robert A Fox, and Joseph Salmons. Vowel change across three age groups of speakers in three regional varieties of american english. *Journal of phonetics*, 39(4):683–693, 2011.

T Florian Jaeger and Esteban Buz. Signal reduction and linguistic encoding. *Handbook of psycholinguistics. Wiley-Blackwell*, 2016.

Keith Johnson. The auditory/perceptual basis for speech segmentation. *Working papers in linguistics-Ohio State University Department of Linguistics*, pages 101–113, 1997.

Keith Johnson. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of phonetics*, 34(4):485–499, 2006. doi: 10.1016/j.wocn.2005.08.004.

Keith Johnson. Decisions and mechanisms in exemplar-based phonology. *Experimental approaches to phonology*, pages 25–40, 2007.

Gitit Kavé and Shiri Nussbaum. Characteristics of noun retrieval in picture descriptions across the adult lifespan. *Aphasiology*, 26(10):1238–1249, 2012. doi: 10.1080/02687038.2012.681767.

Emmanuel Keuleers, Michaël Stevens, Paweł Mandera, and M. Brysbaert. Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The Quarterly Journal of Experimental Psychology*, (8):1665–1692, 2015.

Chin-Hui Lee. On robust linear prediction of speech. *IEEE [Transactions on Acoustics, Speech, and Signal Processing]*, 36(5):642–650, 1987.

Björn Lindblom. Spectrographic study of vowel reduction. *The journal of the Acoustical society of America*, 35(11):1773–1781, 1963. doi: 10.1121/1.2142410.

Björn Lindblom. Articulatory activity in vowels. *The Journal of the Acoustical Society of America*, 36(5):1038–1038, 1964. doi: 10.1121/1.2143314.

Björn Lindblom. Economy of speech gestures. In *The production of speech*, pages 217–245. Springer, 1983. doi: 10.1007/978-1-4613-8202-7$_1$0.

Sue Ellen Linville. The sound of senescence. *Journal of voice*, 10(2):190–200, 1996. doi: 10.1016/s0892-1997(96)80046-4.

Sue Ellen Linville. *Vocal aging*. Singular Thomson Learning, 2001.

Julie M Liss, Gary Weismer, and John C Rosenbek. Selected acoustic characteristics of speech production in very old males. *Journal of Gerontology*, 45(2):P35–P45, 1990. Date revised - 19900901; Last updated - 2017-09-25; SubjectsTermNotLitGenreText - Speech Characteristics 8142 8142PS5 5967 5967PS5 9045 9045PS5 1720 1720PS5; 5276 5276PS5.

Daniel R. McCloy. *phonR: tools for phoneticians and phonologists*, 2016. R package version 1.0-7.

Stephan C Meylan and Susanne Gahl. The divergent lexicon: Lexical overlap decreases with age in a large corpus of conversational speech. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, pages 1006–1011, 2014.

Cornelia Moers, Antje Meyer, and Esther Janse. Effects of word frequency and transitional probability on word reading durations of younger and older speakers. *Language and Speech*, page 0023830916649215, 2016.

Seung-Jae Moon and Björn Lindblom. Interaction between duration, context, and speaking style in english stressed vowels. *The Journal of the Acoustical society of America*, 96(1):40–55, 1994. doi: 10.1121/1.410492.

Richard J Morris and WS Brown. Age-related differences in speech variability among women. *Journal of Communication Disorders*, 27(1):49–64, 1994. Date revised - 19941001; Last updated - 2018-01-16; SubjectsTermNotLitGenreText - Age Differences 225 225PS5; Human Females 4007 4007PS5; Speech Characteristics 8142 8142PS5 5967 5967PS5 9045 9045PS5 1720 1720PS5.

Fermín Moscoso del Prado Martín. Vocabulary, grammar, sex, and aging. *Cognitive Science*, 2016. doi: 10.1111/cogs.12367.

Benjamin Munson and Molly Babel. *(to appear) The Phonetics of Sex and Gender*. Routledge, 2016.

Masaki Nishio and Seiji Niimi. Changes in speaking fundamental frequency characteristics with aging. *Folia phoniatrica et logopaedica*, 60(3):120–127, 2008. doi: 10.1159/000118510.

Gordon E Peterson and Harold L Barney. Control methods used in a study of the vowels. *The Journal of the acoustical society of America*, 24(2):175–184, 1952.

Janet Pierrehumbert. Word-specific phonetics. *Laboratory phonology*, 7:101–139, 2002. doi: 10.1515/9783110197105.101.

Jenny Price. At the crossroads: where age meets accent in the f1/f2 vowel space. In P. Warren and C.I. Watson, editors, *Proceedings of the 11th Australian International Conference on Speech Science and Technology*. Australian Speech Science and Technology Association, 2006.

Hugo Quené. Longitudinal trends in speech tempo: The case of queen beatrix. *The Journal of the Acoustical Society of America*, 133(6):EL452–EL457, 2013. doi: 10.1121/1.4802892.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

Lorraine A Ramig. Aging speech: Physiological and sociological aspects. *Language & Communication*, 6(1-2):25–34, 1986. doi: 10.1016/0271-5309(86)90003-0.

Michael Ramscar, Peter Hendrix, Bradley Love, and RH Baayen. Learning is not decline: The mental lexicon as a window into cognition across the lifespan. *The Mental Lexicon*, 8(3):450–481, 2013. doi: 10.1075/bct.80.08ram.

Michael Ramscar, Peter Hendrix, Cyrus Shaoul, Petar Milin, and Harald Baayen. The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in cognitive science*, 6(1):5–42, 2014.

Michael P Rastatter and Richard D Jacques. Formant frequency structure of the aging male and female vocal tract. *Folia Phoniatrica et Logopaedica*, 42 (6):312–319, 1990. doi: 10.1159/000266088.

Michael P Rastatter, Richard A McGuire, Joseph Kalinowski, and Andrew Stuart. Formant frequency characteristics of elderly speakers in contextual speech. *Folia Phoniatrica et Logopaedica*, 49(1):1–8, 1997. doi: 10.1159/000266431.

Ulrich Reubold and Jonathan Harrington. *Disassociating the effects of age from phonetic change: a longitudinal study of formant frequencies*, pages 9–37. John Benjamins Publishing Company, 2015. doi: 10.1075/impact.37.02reu.

Ulrich Reubold, Jonathan Harrington, and Felicitas Kleber. Vocal aging effects on f0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication*, 52(7):638–651, 2010. doi: 10.1016/j.specom.2010.02.012.

John Rickford and Mackenzie Price. Girlz ii women: Age-grading, language change and stylistic variation. *Journal of Sociolinguistics*, 17(2):143–179, 2013. doi: 10.1111/josl.12017.

Gillian Sankoff. *Longitudinal studies*, pages 261–79. Oxford University Press, Oxford, New York, 2013. doi: 10.1093/oxfordhb/9780199744084.013.0013.

Robert T Sataloff, Deborah Caputo Rosen, Mary Hawkshaw, and Joseph R Spiegel. The aging adult voice. *Journal of Voice*, 11(2):156–160, 1997. doi: 10.1016/s0892-1997(97)80072-0.

Susanne Schötz and Christian Müller. *A Study of Acoustic Correlates of Speaker Age*, pages 1–9. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. doi: 10.1007/978-3-540-74122-0$_1$.

Gail P Scukanec, Linda Petrosino, and Kevin Squibb. Formant frequency characteristics of children, young adult, and aged female speakers. *Perceptual and motor skills*, 73(1):203–208, 1991. doi: 10.2466/pms.1991.73.1.203.

Mark Selent. *Acoustic and Respiratory Measures as a Function of Age in the Male Voice*. PhD thesis, Cleveland State University, 2014.

Scott Seyfarth. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133(1):140–155, 2014. doi: 10.1016/j.cognition.2014.06.013.

Rajka Smiljanić and Ann R Bradlow. Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and linguistics compass*, 3(1):236–264, 2009. doi: 10.1111/j.1749-818x.2008.00112.x.

RJJH Son, Louis CW Pols, et al. How efficient is speech? *Proceedings (Instituut voor Fonetische Wetenschappen, Universiteit van Amsterdam)*, 25:171–184, 2003.

Rob JJH van Son and Louis CW Pols. Information structure and efficiency in speech production. In *Proceedings of Eurospeech 2003*, pages 769–772, 2003.

Elaine T Stathopoulos, Jessica E Huber, and Joan E Sussman. Changes in acoustic characteristics of the voice across the life span: measures from individuals 4–93 years of age. *Journal of Speech, Language, and Hearing Research*, 54(4):1011–1021, 2011. doi: 10.1044/1092-4388(2010/10-0036).

Kenneth N Stevens and Arthur S House. Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech & Hearing Research*, 1963. doi: 10.1044/jshr.0602.111.

Kenneth N Stevens, Arthur S House, and Allan P Paul. Acoustical description of syllabic nuclei: An interpretation in terms of a dynamic model of articulation. *The Journal of the Acoustical Society of America*, 40(1):123–132, 1966.

Erik R Thomas. Sociophonetics. *The handbook of language variation and change*, pages 108–127, 2013.

Fabian Tomaschek, Martijn Wieling, Denis Arnold, and R Harald Baayen. Word frequency, vowel length and vowel quality in speech production: an ema study of the importance of experience. In *INTERSPEECH*, pages 1302–1306, 2013.

Fabian Tomaschek, Benjamin V Tucker, Martijn Wieling, and R Harald Baayen. Vowel articulation affected by word frequency. In *Proceedings of the 10th International Seminar on Speech Production*, pages 425–428, 2014.

Fabian Tomaschek, Denis Arnold, Franziska Bröker, and R Harald Baayen. Lexical frequency co-determines the speed-curvature relation in articulation. *Journal of Phonetics*, 68:103–116, 2018a.

Fabian Tomaschek, Benjamin V Tucker, Matteo Fasiolo, and R Harald Baayen. Practice makes perfect: The consequences of lexical proficiency for articulation. *Linguistics Vanguard*, 4(s2), 2018b. doi: doi:10.1515/lingvan-2017-0018.

Peter Torre and Jessica A Barlow. Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders*, 42(5):324–333, 2009. doi: 10.1016/j.jcomdis.2009.03.001.

Yuichi Ueda, Tomoya Hamakawa, Tadashi Sakata, Syota Hario, and Akira Watanabe. A real-time formant tracker based on the inverse filter control method. *Acoustical science and technology*, 28(4):271–274, 2007.

J.P.H. van Santen. Contextual effects on vowel duration. *Speech Communication*, 11(6):513–546, 1992. doi: 10.1016/0167-6393(92)90027-5.

Irma M Verdonck-de Leeuw and Hans F Mahieu. Vocal aging and the impact on daily life: a longitudinal study. *Journal of Voice*, 18(2):193–202, 2004. doi: 10.1016/j.jvoice.2003.10.002.

Suzanne Evans Wagner. Age grading in sociolinguistic theory. *Language and Linguistics Compass*, 6(6):371–382, 2012. doi: 10.1002/lnc3.343.

Catherine I Watson and Zoe E Evans. Sound change or experimental artifact?: A study on the impact of data preparation on measuring sound change. In *The proceedings of 16th Australasian International Conference on Speech Science and Technology SST 2016*, pages 261–264, 2016.

Catherine I Watson and Jonathan Harrington. Acoustic evidence for dynamic formant trajectories in australian english vowels. *[Journal of the Acoustical Society of America]*, 106(1):458–468, 1999.

Robert L Weide. The CMU pronouncing dictionary. *URL: http://www. speech. cs. cmu. edu/cgibin/cmudict*, 1998.

Melanie Weirich and Adrian P Simpson. Differences in acoustic vowel space and the perception of speech tempo. *Journal of Phonetics*, 43:1–10, 2014. doi: 10.1016/j.wocn.2014.01.001.

Martijn Wieling. Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between l1 and l2 speakers of english. *Journal of Phonetics*, 70:86 –116, 2018.

S. N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73:3–36, 2011.

S. N. Wood. *Generalized Additive Models*. Chapman & Hall/CRC, New York, 2017.

Simon N. Wood, N Pya, and B. Saefken. Smoothing parameter and model selection for general smooth models. *arXiv*, 2016.

An Xue, Jack Jiang, Emily Lin, Raymond Glassenberg, and Peter B Mueller. Age-related changes in human vocal tract configurations and the effects on speakers' vowel formant frequencies: a pilot study. *Logopedics Phoniatrics Vocology*, 24(3):132–137, 1999. doi: doi.org/10.1080/140154399435084.

Steve An Xue and Dimitar Deliyski. Effects of aging on selected acoustic voice parameters: Preliminary normative data and educational implications. *Educational Gerontology*, 27(2):159–168, 2001. doi: 10.1080/03601270151075561.

Steve An Xue and Grace Jianping Hao. Changes in the human vocal tract due to aging and the acoustic correlates of speech productiona pilot study. *Journal of Speech, Language, and Hearing Research*, 46(3):689–701, 2003. doi: 10.1044/1092-4388(2003/054).

Jiahong Yuan and Mark Liberman. Speaker identification on the scotus corpus. In *Proceedings of the Eighth International Conference on Acoustics, Speech, and Signal Processing*, pages 5687–5690, 2008.

	edf	std.dev	lower	upper	Ref.df	Chi.sq	p-value
s(Speaker,Formant)	14.7477	0.1696	0.1133	0.2538	18.0000	5722.5526	0.0000
s(Speaker,Duration,Formant)	8.0419	0.0343	0.0143	0.0822	20.0000	1800.3561	0.3496
s(Speaker,Age)	8.0787	0.0442	0.0260	0.0749	10.0000	124.1549	0.0001
s(Preceding phone,Formant)	57.7277	0.2693	0.2212	0.3280	68.0000	2346.7130	0.0000
s(Following phone,Formant)	52.1603	0.2085	0.1663	0.2615	70.0000	2755.7851	0.0000

Table 4: Summary of the generalized additive mixed model of first and second formants, hand-checked measurements, smooth terms, hand checked measurements

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	7.4969	0.3456	21.6924	0.0000
Sex (male)	-0.9528	0.3324	-2.8668	0.0041
Formant *F2*	1.7468	0.4887	3.5742	0.0004
Vowel æ	0.0849	0.2446	0.3469	0.7287
ɛ	-1.0045	0.2681	-3.7464	0.0002
i	-5.4100	0.2456	-22.0312	0.0000
ɪ	-4.4189	0.2524	-17.5043	0.0000
ɔ	-1.9902	0.2355	-8.4492	0.0000
u	-5.0638	0.2542	-19.9190	0.0000
ʊ	-5.2313	0.4685	-11.1653	0.0000
ʌ	-0.1542	0.2601	-0.5929	0.5532
Duration	0.1352	0.1124	1.2025	0.2292
Age	0.0105	0.0538	0.1950	0.8454
Sex (male):Formant *F2*	0.3108	0.4700	0.6612	0.5085
Sex (male):Vowel æ	-0.0603	0.1260	-0.4786	0.6322
Sex (male): ɛ	0.0265	0.1298	0.2039	0.8385
Sex (male): i	-0.0431	0.1264	-0.3413	0.7329
Sex (male): ɪ	0.0447	0.1262	0.3539	0.7234
Sex (male): ɔ	-0.2464	0.1246	-1.9783	0.0479
Sex (male): u	-0.1607	0.1342	-1.1976	0.2311
Sex (male): ʊ	0.0194	0.1887	0.1030	0.9180
Sex (male): ʌ	-0.2059	0.1300	-1.5845	0.1131
Formant *F2*:Vowel æ	3.0517	0.3459	8.8236	0.0000
Formant *F2*: ɛ	5.1802	0.3790	13.6670	0.0000
Formant *F2*: i	12.0231	0.3471	34.6350	0.0000
Formant *F2*: ɪ	9.1739	0.3568	25.7112	0.0000
Formant *F2*: ɔ	0.3635	0.3331	1.0913	0.2752
Formant *F2*: u	7.1459	0.3595	19.8799	0.0000
Formant *F2*: ʊ	1.7595	0.6625	2.6560	0.0079
Formant *F2*: ʌ	1.2834	0.3678	3.4895	0.0005

	Estimate	Std. Error	z value	Pr(>|z|)
Sex (male):Formant *F2*:Vowel æ	-0.0471	0.1781	-0.2645	0.7914
Sex (male):Formant *F2*: ɛ	0.0063	0.1835	0.0345	0.9725
Sex (male):Formant *F2*: i	-0.0921	0.1786	-0.5158	0.6060
Sex (male):Formant *F2*: ɪ	-0.1460	0.1784	-0.8185	0.4131
Sex (male):Formant *F2*: ɔ	-0.2947	0.1761	-1.6732	0.0943
Sex (male):Formant *F2*: u	-0.2675	0.1898	-1.4095	0.1587
Sex (male):Formant *F2*: ʊ	-0.3706	0.2668	-1.3890	0.1648
Sex (male):Formant *F2*: ʌ	0.0650	0.1838	0.3537	0.7236
Formant *F2*:Vowel æ:Duration	0.6522	0.1462	4.4626	0.0000
Formant *F2*: ɛ:Duration	1.0330	0.1538	6.7155	0.0000
Formant *F2*: i:Duration	2.7140	0.1462	18.5579	0.0000
Formant *F2*: ɪ:Duration	1.9947	0.1464	13.6291	0.0000
Formant *F2*: ɔ:Duration	-0.0222	0.1408	-0.1577	0.8747
Formant *F2*: u:Duration	1.4147	0.1492	9.4821	0.0000
Formant *F2*: ʊ:Duration	0.1040	0.2331	0.4463	0.6554
Formant *F2*: ʌ:Duration	0.2940	0.1527	1.9247	0.0543
Formant *F2*:Vowel æ:Age	0.0301	0.0795	0.3786	0.7050
Formant *F2*: ɛ:Age	0.1176	0.0797	1.4748	0.1403
Formant *F2*: i:Age	0.2591	0.0794	3.2633	0.0011
Formant *F2*: ɪ:Age	0.1663	0.0779	2.1352	0.0327
Formant *F2*: ɔ:Age	0.1085	0.0781	1.3898	0.1646
Formant *F2*: u:Age	0.1879	0.0825	2.2785	0.0227
Formant *F2*: ʊ:Age	0.2605	0.1066	2.4429	0.0146
Formant *F2*: ʌ:Age	0.0578	0.0814	0.7103	0.4775

Table S1: The GAM model of first and second formants (parametric terms), based on the full dataset

	edf	std.dev	lower	upper	Ref.df	Chi.sq	p-value
s(Speaker,Formant)	16.3768	0.4476	0.3081	0.6502	18.0000	549357.3629	0.0000
s(Speaker,Duration,Formant)	18.8410	0.2049	0.1458	0.2879	20.0000	668147.3658	0.0000
s(Speaker,Age)	8.5620	0.0539	0.0324	0.0898	10.0000	852.1585	0.0000
s(Preceding phone,Formant)	57.0161	0.2694	0.2215	0.3276	68.0000	2520.3468	0.0000
s(Following phone,Formant)	48.4370	0.1629	0.1291	0.2055	72.0000	1833.8183	0.0000

Table S2: The The generalized additive mixed model of first and second formants (smooth terms)