



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

The early influence of phonology on a phonetic change

Citation for published version:

Fruehwald, J 2016, 'The early influence of phonology on a phonetic change' *Language*, vol. 92, no. 2, pp. 376-410.

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Language

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



THE EARLY INFLUENCE OF PHONOLOGY ON A PHONETIC CHANGE

Josef Fruehwald

Affiliation: University of Edinburgh

e-mail address: josef.frueh@ed.ac.uk

mailing address:

Linguistics and English Language

The University of Edinburgh

Dugald Stewart Building

3 Charles Street

Edinburgh, EH8 9AD

United Kingdom

THE EARLY INFLUENCE OF PHONOLOGY ON A PHONETIC CHANGE

ABSTRACT: The conventional wisdom regarding the diachronic process whereby phonetic phenomena become phonologized appears to be the “error accumulation” model, so called by Baker, Archangeli, and Mielke (2011). Under this model, biases in the phonetic context result in production or perception errors, which are misapprehended by listeners as target productions, and over time accumulate in new target productions. In this paper, I explore the predictions of the hypocorrection model for one phonetic change (pre-voiceless /ay/ raising) in detail. I argue that properties of the phonetic context under-predict and mischaracterize the contextual conditioning on this phonetic change. Rather, it appears that categorical, phonological conditioning is present from the very onset of this change.*

Keywords– phonologization, hypocorrection, phonetic change, Canadian raising, sociolinguistics

* Thanks to Bob Ladd, James Kirby, Ricardo Bermúdez-Otero and Natasha Warner for their comments on this paper, as well as to William Labov, Eugene Buckley, Charles Yang, Tony Kroch, Joel Wallenberg, Meredith Tamminga, Márton Sóskuthy, Michael Becker, Karthik Durvasula, Andrew Nevins and Alan Yu for their comments on earlier versions of this work. I would also like to thank audiences at the NWAV, NELS, and mFiL conferences, the Symposium on Historical Phonology, as well as audiences at Newcastle and York Universities. Additional thanks are due to the anonymous reviewers at Language for their invaluable comments. All errors are my own.

1. INTRODUCTION. As a general model of sound change, hypocorrection, proposed by Ohala (1981), appears to be a solution to a number of problems in phonology, phonetics, and sound change, including the Actuation Problem, the Incrementation Problem, the naturalness of sound change, and the naturalness of phonological processes. The Actuation Problem, as originally formulated by Weinreich, Labov, and Herzog (1968), poses the question why a particular sound change occurred where it did and when it did. The Incrementation Problem, as formulated by Labov (2001), poses the question about how a language change can continue in the same direction over multiple generations. Blevins (2008) defines “natural sound patterns” as “those grounded in physical properties of speech.” One explanation for the preponderance of “natural” sound patterns in this sense would be that phonological markedness constraints are themselves grounded in “speakers’ shared knowledge of the factors that affect speech communication by impeding articulation, perception, or lexical access” (Hayes and Steriade 2004). Of course, the Evolutionary Phonology explanation that Blevins (2004) expounds places the explanatory power for natural sound systems in the fact that some kinds of sound changes are more common than others due to their phonetic grounding, a point further elaborated by Garrett and Johnson (2013).

The canonical illustration of hypocorrection is /u/ fronting from Ohala (1981), which can be summarized as follows. When adjacent to a coronal, like /t/, a /u/ may be perceived as further front, perhaps as [ʉ]. This could be either due to physical coarticulation with the tongue body moving towards a fronter target for /u/ than usual, or due to the acoustic effect of a [t] closure on the formant structure of adjacent /u/, perceptually fronting it. This difference between physical coarticulation and acoustic warping is largely immaterial to the hypocorrection model, as long as the listener’s percept is different from the speaker’s intended production. Hypocorrection occurs when a listener fails to take into account the contextual effects of the speaker’s production, and instead reconstructs the speaker’s intended production target as [ʉ]. Over the course of many interactions between speakers and listeners, with listeners hypocorrecting, a critical mass of speakers in a speech community may have an underlying distribution of /ʉ/ when adjacent to coronals, and /u/ otherwise, at which point it might be said that the language has changed.

The way in which hypocorrection resolves the incrementation and naturalness problems is straightforward. The fundamental mechanism of hypocorrection rests in the physical world (the acoustic properties of the human vocal tract) and the finite precision of human motor planning.

There is a natural and persistent pressure in this model to front a [u] target to a [ʊ] production. The incrementation problem (Why does sound change progress in the same direction over many generations?) is resolved, because both the human articulatory and perceptual systems are remaining constant across the relevant time periods, ensuring a constant bias. An explanation of the naturalness of phonological processes can also be found in hypocorrection. In the case of /u/ fronting, the new distribution of [u] and [ʊ] could be described as a phonological rule, like (1).

(1) $u \rightarrow [-\text{back}] / [\text{COR}]$

If it is assumed that most of the contents of phonological grammars are either the end products of sound changes or the further elaboration and modification of those end products (Blevins 2004; Bermúdez-Otero 2007) and that some sound changes are more likely to happen than others because of their articulatory and perceptual grounding (Garrett and Johnson 2013), then it would follow that most phonological processes would be natural, due to their historical origins. The actuation problem is not, however, resolved by the hypocorrection model, as was nicely demonstrated by Baker, Archangeli, and Mielke (2011). Rather than just asking “What actuated a sound change?” the actuation problem really asks “Why did this change occur now, in this dialect, and never before, and not in the neighboring dialects?” Since the motivating factors under hypocorrection are natural and persistent, thus present in all dialects at all times, no good answer to the actuation problem is immediately available. The point that hypocorrection, on its own, cannot account for the actuation of sound changes was made in Ohala and Ohala (1991). However, Yu (2013) has found inter-listener differences with respect to how much they hypocorrect, and suggested that these differences in “cognitive style” may covary with speakers’ embedding in social networks. If the hypocorrectors in the speech community are properly placed, socially, then the hypocorrection change may spread throughout that speech community. Baker, Archangeli, and Mielke (2011) suggest that this alignment of the phonetic motivation of a sound change with social dimensions that make its propagation possible might be a relatively rare occurrence. These two lines of inquiry may go some ways towards resolving the question of why phonetically motivated sound changes do not all inevitably happen in all languages.

The hypocorrection model is supported by many reasonable assumptions. Experimental work has shown that listeners do hypocorrect, and the way that they do mirrors attested historical changes (Ohala 1990; Harrington, Kleber, and Reubold 2007; Yu 2013 among others). In addition, the outcomes of various simulations of hypocorrection appear similar to the outcomes

of attested language changes given similar inputs (Pierrehumbert 2001; Wedel 2007; Garrett and Johnson 2013 among others). One missing strain of evidence for the hypocorrection model of language change is data from actual language change in progress, but this will soon be changing. Variationist sociolinguistics, historically the subfield devoted to the study of language change in progress, is increasing the volume, quality, and time depth of data available to researchers to address questions such as these. Corpora such as the Buckeye Corpus (Pitt et al. 2007), the Origins of New Zealand English (ONZE) corpus (Gordon, Maclagan, and Hay 2007), and the Philadelphia Neighborhood Corpus (Labov and Rosenfelder 2011) are such examples.

In this paper, I investigate the role that hypocorrection plays in a conditioned sound change in the Philadelphia Neighborhood Corpus: the raising of /ay/ before voiceless consonants. This logic of this investigation attempts to evaluate the prediction of the hypocorrection model that phonetic changes are necessarily circumscribed by other phonetic properties of speech. In the hypothetical example of /ut/ fronting, the phonetic precursor to /u/ fronting was the fronting of the [u] percept, either due to physical coarticulation or acoustic warping from adjacent coronals on the [u] formant structure. Before /u/ fully fronted to [ʊ], this precursor should have been detectable in the speech community.

Importantly, the hypocorrection model predicts that the rate of a hypocorrective change should be proportional to the size of the phonetic precursor. I'll be defining the precursor "size" in terms of the difference between a speaker's intended phonetic production, and listener's percept. In practical terms of this illustration with /u/, that will be the effect a following [t] has on F2, agnostic to whether that effect is due to a fronter [u] articulation, or the acoustic effect an alveolar closure has on the formant structure. To illustrate how precursor size ought to influence the rate of change, suppose in the language, /u/ appeared in two different coronal contexts, pre-alveolar [t] and pre-retroflex [ʈ]. By virtue of its more posterior place of articulation, [ʈ] would probably exert a weaker coarticulatory/perceptual pressure to front than [t]. As such, when a listener hypocorrects, and incorporates error into their own representation of /u/, that error would be less for /uʈ/ (say [ʊʈ]) than for /ut/ (say [ʊt]). To put illustrative numbers on it, the effect of a following [t] would have a +0.5 fronting bias on [u] and a following [ʈ] would have a +0.3 fronting bias. Over time, and multiple interactions, this would have two results:

1. A slower rate of /uʈ/ fronting (+0.3 per time step) than /ut/ fronting (+0.5 per time step).
2. A less fronted realization for /uʈ/ than /ut/.

Both of the following would be unexpected under the straightforward hypocorrection account, and it would require further elaboration to account for them.

1. An identical rate of fronting for both /u̟/ and /u/.
2. Fronting of /u̟/, but not /u/.

In addition, if there were a (morpho)phonological process which turned /t/ into /k/ in some contexts, it would be expected that the coarticulatory/perceptual effect of /u/ fronting to be non-existent before /k/. If a speaker were to hypocorrect, they would update their representation with the error [u̟t], but there would be no such error in the context of [uk]. Again, to put illustrative numbers on it, the fronting bias for a following [t] would again be +0.5, but for a following [k] it would be 0. The result would be:

1. /u̟/ would front at the rate of +0.5 per time step, and /uk/ would not front at all.

If it were observed that /uk/ were participating in the fronting change to a similar degree as /u̟/, then that would be evidence that something a bit more elaborated than hypocorrection is at work. Lexical analogy or paradigm uniformity might be appealed to explain /uk/ fronting, two factors well outside of strict phonetic conditioning.

Figure 1 illustrates how the differential fronting biases of [t], [t̟] and [k] would accumulate over time, resulting in different rates of change for /u/ in these contexts, and in differently fronted realizations of /u/ at the end point.

<INSERT *Figure 1* ABOUT HERE>

These rate of change effects are predicted by the error-accumulation model of hypocorrection, as Baker, Archangeli, and Mielke (2011) called it, or the production-perception feedback loop model, as Wedel (2006) called it. However, as pointed out by a reviewer, there may be ways in which hypocorrection influences phonetic change other than the way modeled in these error-accumulation approaches. For example, it could be that contexts that exert maximal bias are more likely to actuate a sound change, or are more likely to do so earlier. This would result in a lag between when the most biasing context begins to change and when the least favoring context does. Alternatively, the change may begin in all possible contexts at the same time and progress at the same rates, but go on for longer in the most biasing context. These two scenarios still both predict different diachronic trends for our example vowel /u/, either through delayed actuation

early in the change or through extended change later. The outcome at the end points is also the same: the most strongly biasing context has the most extremely fronted realization. For the remainder of the paper, I will be implicitly couching my analysis against the predictions of the error-accumulation model, largely because of its intuitive appeal and popularity in the literature, assuming that different bias strengths ought to result in different rates of change. However, if one of these other two hypocorrection models (actuation modulation and change extension) were assumed, they would not fundamentally change the interpretation of the results below.

The example I examine in this paper is directly analogous to the illustrative examples just given. The data is drawn from the Philadelphia Neighborhood Corpus of Ling 560 studies (“the PNC” from here on). I look at the raising of /ay/ from a low nucleus to a mid nucleus before voiceless consonants in Philadelphia (“Canadian” Raising), specifically with how this diachronic change interacts with /t/ and /d/ flapping. I establish that the phonetic precursors which have been suggested for /ay/ raising are either neutralized or strongly mitigated in the flapping context, but despite this fact, the degree to which /ay/ raises or remains low in these flapping contexts appears more or less identical to how it raises or remains low in “faithful,” non-flapping contexts at all time points throughout the change. It is the underlying phonological context which best predicts the degree of /ay/ raising, not the phonetic properties of that context.

This case study is an example of a phonetic change progressing at a rate which is disproportionate to the size of the phonetic precursors. This should be surprising under the hypocorrection model of phonetic change. I argue that the selection of contexts to undergo /ay/ raising is categorical and phonological. Before concluding, I will examine one alternative explanation (lexical analogy), and will show that insofar as it can be quantitatively operationalized, it does not account for the observed patterns.

2. METHODS

2.1. DATA USED. Unless otherwise indicated, all vowel formant data in this paper is drawn from the Philadelphia Neighborhood Corpus of Ling 560 studies (Labov and Rosenfelder 2011), henceforth the PNC. All word frequency data is drawn from SUBTLEX_{US} (Brysbaert and New 2009).

The PNC consists of sociolinguistic interviews carried out throughout Philadelphia by graduate students as part of their course work for Ling 560, *Researching the Speech Community*, at the University of Pennsylvania. The course has run from 1973 to 2012. Every year, two to three research groups would choose a city block in Philadelphia and attempt to interview as many residents on that block as possible. At the time of writing, 397 of these interviews have been digitized, and approximately 232 hours have been transcribed and force-aligned using the FAVE-suite (Rosenfelder et al. 2014).

In this paper, I will be focusing on the 326 white, Philadelphia-born speakers in the corpus (198 hours of transcribed speech). The earliest date of birth of a speaker in this subsample is 1889, and the most recent is 1998. Figure 2 displays the distribution of ages across each year of the study.

<INSERT Figure 2 ABOUT HERE>

Point estimates of F1 and F2 were automatically extracted from all vowels with primary lexical stress in this subsample using the Bayesian formant tracking technique from the FAVE-suite (see Labov, Rosenfelder, and Fruehwald (2013) for a more complete description), resulting in 629,069 vowel formant estimates. While the FAVE-suite estimates F1 and F2 at 1/3 of the vowel's duration for most vowel classes, they are estimated at F1 maximum for /ay/ and /ey/. Evanini (2009) found that when comparing this automated method of formant estimation to the manual measurements from the Atlas of North American English (Labov, Ash, and Boberg 2006), there was a mean absolute difference of about 50 Hz on F1 and 90 Hz on F2 between the two. Within each speaker, F1 and F2 were converted to z-scores (i.e. Lobanov normalization). Adank, Smits, and Hout (2004) found that z-score normalization was the most effective at eliminating physiological differences while preserving social differences, and Rathcke and

Stuart-Smith (2014) found that it was most effective at mitigating artifacts caused by poor signal-to-noise ratios and more extreme spectral tilt sometimes found in archival recordings.

2.2. ANALYSIS PROCEDURE. All tokens of /ay/ were extracted from the speakers' full data sets, excluding proper names for the sake of speakers' anonymity. This resulted in 84,713 /ay/ measurements. Of particular interest here is the F1 of the /ay/ nucleus, which will be taken as the indicator of vowel height, as well as other contextual information about the /ay/ token, including the voicing of the following segment, which is all included in the output of FAVE-extract.

Section 3 is devoted to establishing the broad facts of /ay/ raising so that the details of how it interacts with flapping can be properly understood. In the first instance, establishing that pre-voiceless /ay/ raising is truly a language change in progress will involve the triangulation between descriptions of the Philadelphian dialect from the 1940s, 1970s and 2000s, as well as the comparison of these descriptions to the data available in the PNC. I then carry out analyses of the distributional properties of /ay/, and fit both non-linear and linear mixed effects models in order to establish the dynamics of this change. After establishing the broad facts, §4 investigates the phonetic conditioning of /ay/ raising, specifically when followed by flaps. After identifying tokens of /ay/ which unambiguously appear before “faithful” /t/ and /d/, where the voicing distinction is not neutralized, and tokens which appear before flapped realizations of /t/ and /d/, I explore whether the previously proposed phonetic precursors for /ay/ raising are neutralized in the flapping contexts. The two precursors explored here are pre-voiceless shortening, originally proposed as the phonetic precursor for Canadian Raising by Joos (1942), and pre-voiceless offglide peripheralization, proposed by Moreton (2004). After establishing that these precursors are either neutralized or minimized in flapping contexts, I turn to non-linear Bayesian Modeling to establish whether at any point across the 20th century, /ay/ preceding flaps patterned according to the phonetic properties of the flapping context or according to the underlying phonological identity of the flap.

3. /AY/ RAISING: A PHONETIC CHANGE IN PROGRESS. As outlined in the introduction, the key contribution of this study is investigating the role of phonetic conditioning on a sound change in progress. It is therefore crucial that I first establish that pre-voiceless /ay/ raising is, in fact, a change in progress and that it is progressing as a classic Neogrammarian sound change. There are alternate routes by which Philadelphia could have arrived at an /ay/ raising system which would defeat the effectiveness of using it as an example of a conditioned sound change. For instance, pre-voiceless /ay/ raising could be historical residue of the Great Vowel Shift, whereby /ay/ used to be uniformly realized as [əi], but lowered to [ai] except before voiceless consonants. This is called the “failure to lower” hypothesis by Moreton and Thomas (2007), and was appealed to explain the distribution of /ay/ variants in Martha’s Vineyard by Labov (1963). If Philadelphia’s /ay/ raising system was due to a failure to lower pre-voiceless /ay/, then the phonetic conditioning that gave rise to it is lost to history, and this investigation would no longer concern a true language change *in progress*. Alternatively, pre-voiceless /ay/ raising could have entered Philadelphia via dialect contact with a region with a fully developed /ay/ raising system. Again, if this were the case, then Philadelphians’ (non)conformity to the predictions of purely phonetic conditioning would be uninformative, since the relevant historical facts would have to be investigated in the originating dialect Philadelphia had contact with. Finally, even if pre-voiceless /ay/ raising is a change in progress captured within the PNC data and is truly endogenous to Philadelphia, it might not be a classic Neogrammarian sound change, progressing via gradual phonetic adjustments of the vowel nucleus. Instead, it could progress via the competition of categorical variants [ai] and [ʌi], with the frequency of [ʌi] gradually increasing over time. This final possibility might not rule out /ay/ raising’s utility for exploring phonetic conditioning of sound change in progress, but our analysis methods would need to be adjusted to take into account categorical variation.

This section is devoted to the exploration of these issues, so that it can be assured that the detailed analysis of /ay/ raising in flapping contexts sits on a solid foundation. I begin by combining a review of the earliest description of the Philadelphia dialect, more contemporary sociolinguistic research on Philadelphia, and the data available in the PNC. This is followed by more detailed quantitative analysis of the distributional properties of /ay/, concluding with a

broad characterization of the dynamics of the change. Fortunately, none of the hypothetical scenarios that would limit /ay/ raising's illustrativeness are supported by the available evidence.

3.1. A CHANGE IN PROGRESS. In his (1944) description of the Philadelphian dialect, Tucker said

Both the [aɪ]-type diphthong and the [aʊ]-type diphthong exist in only one quality, whereas in most American dialects the first element is shortened and modified in quality before a voiceless consonant - the precise sounds vary according to locality. (In my own speech, for example, this short sound, as in *night* or *out*, seems to be identical with the vowel of *but*; contrast with [ɔ] in *ride*, *loud*.) No such distinction is made in the Philadelphia dialect.

It's not clear on which speakers (and importantly, of what generation) Tucker based this statement, but he is very explicit on this point that Philadelphians did not exhibit any raising of pre-voiceless /ay/. His other descriptions of vocalic variation in the same paper are similarly explicit, and concordant with available data, so there is no reason to doubt that his description is accurate in this case.

By the 1970s, the Language Change and Variation project at the University of Pennsylvania found that pre-voiceless /ay/ raising was a vigorous change in progress, led by men (Labov 2001). Subsequent sociolinguistic research on pre-voiceless /ay/ has found that its raising trend has slowed, and that now backer phonetic realizations are associated with masculinity and toughness (Conn 2005; Wagner 2007).

<INSERT Figure 3 ABOUT HERE>

Setting aside the effect of sex, which for this particular sound change is mitigated in the PNC, the diachronic pattern as described beginning in the 1940s, through to the 1970s and late 2000s is corroborated by the PNC data. Figure 3 plots mean normalized F1 of pre-voiced and pre-voiceless /ay/ for each speaker against their date of birth.¹ While Tucker does not specify any demographic information about the speakers he based his description on, the data in Figure 3 is broadly compatible with his description. The difference between *write* and *ride* doesn't seem to be robustly present until approximately 1920. Speakers born in 1920 would have been 24 at the time Tucker published his notes on the Philadelphia dialect, and if the speakers he based his descriptions on were older than that, on average, they would have been born even earlier, and

made even less of a distinction than that. Since that time, pre-voiceless /ay/ has changed to an incredible degree. In Figure 3, horizontal lines representing the average normalized F1 of /a/ and /ʌ/ (neither of which exhibit any notable diachronic trend of their own) have been superimposed on the diachronic trends for these /ay/ allophones. For speakers born before the turn of the 20th century, both allophones had a nucleus slightly lower than /a/, but for speakers born around 1970, the nucleus of pre-voiceless /ay/ has risen to be equivalently high as /ʌ/, while elsewhere it has remained at approximately the same lowness for the full century.

3.2. AN ENDOGENOUS CHANGE. The operating assumption of sociophoneticians when looking at a trend like Figure 3 is that it represents a phonetically gradual change. That is, speakers born in 1900 produced their pre-voiceless /ay/ something like [aɪ], speakers born in 1970 produced their pre-voiceless /ay/ something like [ʌɪ], and speakers born in between produced them in some phonetically intermediate way. As mentioned above, however, there are other ways this change could have occurred, perhaps via dialect contact. For example, there may have been a poorly documented inflow of speakers from Canada (Joos 1942; Chambers 1973) or the Northern United States (Dailey-O’Cain 1997; Vance 1987) into Philadelphia who brought with them a fully developed pattern of Canadian Raising. The trend in Figure 3 would then be due to population replacement, rather than any linguistic change. If not due to population transfer itself, dialect contact and diffusion (Labov 2007) could still be the culprit. Perhaps a raised [ʌɪ] variant entered Philadelphians’ repertoire and gradually spread either through the lexicon, or through increasing frequency of use. Categorical variation in vowel quality is certainly possible. For example, Smith, Durham, and Fortune (2007) found that categorical variation between [ʌu] and [u:] in words like *down* is linked to style shifting in child-caregiver interactions in North-Eastern Scotland. However, insofar as it can be determined from the PNC data, the default sociophonetic interpretation of gradual phonetic change best characterizes the data.

To begin with, there does not appear to be two different populations represented in Figure 3. There is no sudden appearance of fully raised [ʌɪ] speakers, nor a gradual disappearance of [aɪ] speakers. In order to visualize this fact more clearly, I calculated the Cohen’s *d* estimate of the voicing effect size for every speaker. Cohen’s *d* is an estimate of the difference between two distributions, standardized by the variance of the data. Two distributions with a very narrow

range of variation within them will have a larger Cohen's d than two other distributions with the same difference in means, but a broader variance. Since the largest possible difference between the means of two distributions is infinite, the largest possible value of Cohen's d is also ∞ . In this case, the difference between two means was always calculated by subtracting the mean normalized F1 for pre-voiceless /ay/ from that of pre-voiced /ay/. For speakers with a lower pre-voiceless /ay/ than pre-voiced /ay/, this will result in a negative Cohen's d . This difference between the means was then divided by the pooled standard deviation.

If a population with a mature and strong Canadian Raising grammar moved into Philadelphia and replaced the non-raising speakers, a pattern of two horizontal stripes in the Cohen's d estimates would be expected. The top stripe would represent the incoming speakers for whom following voicing has a strong effect on vowel quality, and the bottom stripe would represent the speakers being replaced for whom following voicing does not have a strong effect. Gradually, the bottom stripe would fade out, and the top stripe would fade in. This hypothetical situation is clearly not the case if looking at Figure 4, which plots every speaker's Cohen's d estimate against their date of birth. Rather than two horizontal stripes representing two different populations of speakers, the pattern instead appears to show one population of speakers which is gradually shifting from having a small effect of following voicing on vowel quality to having a very large effect.

<INSERT Figure 4 ABOUT HERE>

Furthermore, as discussed by Fruehwald (2013), most northern dialects from which Philadelphia may have borrowed /ay/ raising from usually exhibit raising before /r/ as well (Vance 1987; Dailey-O'Cain 1997). Philadelphia, on the other hand, only exhibits raising before voiceless consonants. Following /r/ does not condition the change at all, making it even more unlikely that the origins of /ay/ raising in Philadelphia is due to dialect contact.

3.3. A PHONETIC CHANGE. Neither a plot of speakers' means, nor the plot of Cohen's d would be able to distinguish between continuous and categorical variation within speakers. If speakers born around 1940 produced [ɑɪ] 50% of the time, or for 50% of their words, and [ʌɪ] the other 50%, their points in Figure 3 and Figure 4 would look approximately same as if they produced something phonetically intermediate between [ɑɪ] and [ʌɪ] 100% of the time. One way to try to distinguish between categorical and continuous variation is to look at the distributional properties of speakers' data. If, unbeknownst to the researcher, a speaker's pre-voiceless /ay/ productions were drawn from two distributions, one [ɑɪ] and the other [ʌɪ], the standard deviation of this mixed distribution would be greater than the standard deviation of the two distributions in isolation, and the kurtosis would be less. The kurtosis of a distribution can be thought of either as how sharply peaked a distribution is, how thick its tails are, or, as Darlington (1970) argued, how unimodal it is.

<INSERT Figure 5 ABOUT HERE>

Figure 5 plots the kurtosis of every speaker's pre-voiceless and pre-voiced /ay/ distributions against their date of birth. A smaller kurtosis corresponds to a flatter, or more bimodal distribution. All normal distributions have the same kurtosis ($=3$), so a horizontal line at $y=3$ has been drawn for reference. If the diachronic trend of pre-voiceless /ay/ raising was due to the categorical replacement of an [ɑɪ] distribution with an [ʌɪ] distribution, then a dip in kurtosis, reaching its minimum at the change's midpoint where the two distributions would be most evenly mixed, would be expected. From the change's midpoint to its endpoint, kurtosis would begin to increase, reaching its maximum when speakers begin drawing exclusively from an [ʌɪ] distribution. It should also look markedly different from the kurtosis of pre-voiced /ay/ which didn't undergo any considerable change across this time period. This hypothetical kurtosis profile is not observed for pre-voiceless /ay/ raising. Instead, the trend in kurtosis across the 20th century appears to be more or less flat, with a median value of 3.06, suggesting speakers aren't using a mixture of two categorical pronunciations.

<INSERT Figure 6 ABOUT HERE>

Turning now to the standard deviation of speaker's pre-voiceless /ay/ distribution, it would exhibit the opposite profile relative to the change from the kurtosis if speakers were drawing from two distinct distributions. The mixture of two normal distributions will always have a standard deviation greater than the standard deviation of either of the contributing distributions. As such, speakers at the midpoint of the change, where the mixture of [ɑɪ] and [ʌɪ] would be most even, would have a greater standard deviation than the beginning or end of the change, where speakers would only be drawing from one or the other distribution. Looking at Figure 6, it appears as if, again, this hypothetical profile of a categorical shift from [ɑɪ] to [ʌɪ] is not operating. The average standard deviation of pre-voiceless /ay/ appears to be lower than pre-voiced /ay/, probably because it appears in a more restricted set of phonetic environments. This is the opposite of what would be expected if pre-voiceless /ay/ raising was a result of mixing [ɑɪ] and [ʌɪ] pronunciations. Included in Figure 6 is a rolling estimate of the inter-speaker standard deviation of speakers' means, based on a window of 20 years, calculated over 5 year increments. This provides some idea of how closely speakers from any given date of birth cohort are clustered together. At any point in time, the degree of between-speaker variation is always less than the degree of within-speaker variation. This information is less relevant to the nature of individuals' behavior, but goes to show that this change wasn't characterized by factions of undergoers and non-undergoers, but rather by relatively strong cohesion within birth cohorts across the entire speech community.

3.4. DYNAMICS OF THE CHANGE. This section will conclude with a broad analysis of the dynamics of this change and by carrying out some standard analyses for any sound change in progress. To begin with, I'll attempt to justify relying exclusively on Date of Birth to characterize the time dimension of this change. There is also always some concern when relying exclusively on apparent time (Sankoff 2006) (in this case, speakers' date of birth) that some non-trivial degree of lifespan change is being overlooked. Treating speakers as speech time capsules of the era in which they were born is, of course, a strong idealization, and a number of panel and case studies have found that speakers can, in fact, change their speech well past the critical period (Harrington, Palethorpe, and Watson 2000; Sankoff and Blondeau 2007). While the PNC does have an important real-time component to it, it is difficult to disentangle the relationship between age and date of birth in its data. By the simple fact that its data was collected between 1973 and 2012, it is impossible for it to contain data from 20 year olds born in 1890, nor 80 year olds born in 1993. Labov, Rosenfelder, and Fruehwald (2013) argued that date of birth provided the best explanatory power for the data when compared to age or year of study. In an attempt to improve on that argument here, I fit a generalized additive model, estimating speakers' F1 based on a two dimensional tensor product between age and date of birth using the *mgcv* package in R (Wood 2011; Wood 2014). This allows us to estimate a non-linear effect of age, date of birth, and their interaction, on pre-voiceless /ay/. Figure 7 plots the estimated lifespan trend for a number of date of birth cohorts. The prediction lines are clipped to the range of what would have been observable within the PNC. While the estimated lifespan trend for speakers born in, say, 1953, is not perfectly flat, it is still not so extremely different from flat that a gross error will be committed by only taking into account speakers' date of birth when modeling the change in pre-voiceless /ay/.

<INSERT Figure 7 ABOUT HERE>

To conclude, I fit a standard mixed effects model including speakers' date of birth, following voicing, word frequency. Word frequency norms were drawn from SUBTLEX_{US}, which contains word counts compiled from the subtitles of U.S. films and television shows. Brysbaert and New (2009) found that these word frequency norms better accounted for participants' behaviors in

lexical decision tasks than frequency norms from other sources, meaning they have good psychological validity. Specifically, the \log_2 transform of the expected word frequency per 1 million words will be used, centered around the median. The full model tried to predict normalized F1 by the three way interaction of date of birth (centered at 1950 and divided by 10, providing the rate of change per decade), \log_2 of word frequency, and voicing context (voiced vs voiceless), including random intercepts for speakers and words, with a random slope of date of birth by word, and of voicing context by speaker.² The model was fit using the lme4 package, version 1.1-7, in R (Bates et al. 2014), and 95% confidence intervals for the fixed effects were estimated via semiparametric bootstrap replication, using the bootMer function. The fixed effects estimates, 95% confidence intervals and density distribution plots of the bootstrap replications are displayed in Table 1.

<INSERT Table 1 ABOUT HERE>

The fixed effects estimates indicate that there is a main effect of frequency on the F1 of /ay/, but that this effect is not different across the voicing contexts. The effect labeled “Doubling Frequency” corresponds to the effect of doubling frequency on pre-voiced /ay/, and it appears to be reliably different from 0 based on the bootstrap confidence intervals. However, the interaction of “Doubling Frequency \times [-voice]” is not reliably different from 0, meaning the frequency effect is more or less the same between pre-voiced and pre-voiceless /ay/. The raising effect of doubling frequency is approximately equal to one year in date of birth for pre-voiceless /ay/.³ The fact that word frequency affects both pre-voiced and pre-voiceless /ay/ makes sense if it is taken into account that more frequent words are more likely to undergo phonetic reduction, which in the case of these low vowel nuclei would mean lowering F1. However, word frequency does not appear to reliably interact with the rate of pre-voiceless /ay/ raising. The estimated effect size of the three way interaction “Decade \times Frequency \times [-voice]” is exceptionally small, and not reliably different from 0. These results are very similar to what Zellou and Tamminga (2014) found in the PNC for vowel-nasal coarticulation. More frequent words experienced more reduction/coarticulation, but frequency did not interact with changes in nasal-coarticulation. This constitutes a “stationary” effect of word frequency (cf. Hay et al. (2015)).

If frequency is entered into the model in a stepwise fashion, and models are compared using likelihood ratio tests, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), the only improvement to the model is made when it is entered as a main effect, without any interactions (Table 2). It can be tentatively concluded, then, that the effect of word frequency on the change (i.e. its interaction with date of birth) is marginal at best, and perhaps even non-existent.

<INSERT Table 2 ABOUT HERE>

Table 3 displays the random effects' standard deviations and correlations from the full model summarized in Table 3. Examining the standard deviation of [-voice] by speaker, and the residual standard deviation, it can be seen that these values are broadly similar to the estimates of the within and between speaker standard deviations from Figure 6. These are sensible interpretations of these random effects' standard deviations, and the fact they are so similar to the maximum likelihood estimates from Figure 6 is a sign that the model was sensibly fit.

<INSERT Table 3 ABOUT HERE>

3.5. PRELIMINARY CONCLUSIONS ABOUT PRE-VOICELESS /AY/ RAISING. On the basis of historical descriptions of the dialect and careful analysis of the data available in the PNC, a few conclusions can be drawn about the nature of pre-voiceless /ay/ raising with some certainty. First, it is an innovation that occurred in Philadelphia within the 20th century. Triangulating between Tucker's 1942 description and the available PNC data, it is not likely that Philadelphians born before 1900 made a distinction in vowel quality between *ride* and *right*. The nature of the innovation also seems clear. After calculating the Cohen's d effect of voicing on /ay/, it does not appear to be the case that speakers with [ʌɪ] for pre-voiceless /ay/ moved into Philadelphia and replaced the non-raising population, nor does it appear that the speech community was ever divided into a group of speakers who had raising and those who didn't. Looking at the distributional properties of /ay/ (kurtosis and standard deviation), it does not appear that a raised [ʌɪ] was borrowed into the speech community as a new pronunciation variant, nor that there was a period where categorical variation between [ɑɪ] and [ʌɪ], whether lexically conditioned or otherwise, was the norm. Rather, it looks as if the standard sociophonetic assumption is borne out, summarized in (2).

(2) Pre-voiceless /ay/ changed by a gradual shift in the center of its distribution that propagated in a continuous fashion across generational cohorts.

4. /AY/ RAISING AND /T/, /D/ FLAPPING. Of course, "Canadian" Raising is of great interest to phonologists for its opaque interaction with American /t/ and /d/ flapping (Joos 1942; Mielke, Armstrong, and Hume 2003; Idsardi 2006; Pater 2014). Even when the voicing contrast is neutralized (or minimized), /ay/ raising still occurs, so that the distinction between *riding* [ɑɪ] and *writer* [ʌɪ] is maintained in the vowel quality of the preceding /ay/. In contemporary Philadelphia, /ay/ raising does occur before flapped /t/ (this will be demonstrated below) and does not occur before flapped /d/ as a general rule, but there are some lexically conditioned exceptions (Fruehwald, 2008). This lexical diffusion pattern appears to be a separate phenomenon overlaid on top of the general raising pattern (Fruehwald, 2013) and will not be addressed here.

Given that /ay/ raising in Philadelphia appears to be a phonetically gradual innovation, and that it currently interacts opaquely with flapping, the question arises as to whether /ay/ raising before /t/ flaps is phonetically unexpected at all. Perhaps the phonetic properties of the pre-/t/-

flap context are such that raising can be expected there. There are three necessary steps to examine this question. First, tokens of /ay/ that are almost certainly preceding flaps, and faithful /t/ and /d/ need to be identified. Second, the most probable phonetic precursors for /ay/ raising need to be identified. Third, it needs to be determined in which contexts ([faithful, flapped] × [/t/, /d/]) these precursors were present, and at what strength.

4.1. IDENTIFYING FAITHFUL AND FLAPPED /T, D/. For the first step, /ay/ tokens which were most probably preceding faithful and flapped /t/ and /d/ were identified in the PNC. This was done by identifying phonological contexts where /t/ and /d/ are expected to be flaps or stops. Search criteria were defined based on the phone-level alignment of the transcription. The search definition for /ay/ preceding faithful /t/ and /d/ is given in (3) and the search definition for /ay/ preceding flapped /t/ and /d/ is given in (4).

(3) Faithful (stop realizations):

- a) /ay/ is followed by /t/ or /d/.
- b) The /t/ or /d/ is followed by a word boundary.
- c) The /t/ or /d/ is followed by a pause.

(4) Flap:

- a) /ay/ is followed by /t/ or /d/.
- b) The /t/ or /d/ is not followed by a word boundary.
- c) The /t/ or /d/ is followed by an unstressed vowel.

Restricting the definition of faithful /t/ and /d/ to be word final and followed by a pause ensures that no tokens of phrase level flaps will be included. Occasionally, the aligner will mis-label a long final closure as a pause, but for these purposes that is a beneficial error, because if the closure was long enough to be mis-labeled a pause, it was certainly not a flap. The search definition for flaps may be too restrictive, excluding some flapping of /t/ and /d/ that occurs at the phrase level. However, it is better here to be too restrictive than permissive, especially when it can be afforded due to the volume of data available. The resulting numbers of tokens in each context are given in Table 4.

<INSERT Table 4 ABOUT HERE>

There may be some reasonable concern that this rule based definition of flaps, which does not take into account any phonetic properties, may overapply and label tokens as flaps which are not, in fact, flaps. In order to address this concern, I examined all of the /t/-flap tokens and impressionistically coded them on the basis of the spectrogram. If there was a voicing bar, and/or clear formants in the purported flap, it was coded as being a flap. Much like Warner and Tucker (2011) found, there were quite a few tokens of /t/ flaps that were realized as approximant-like (exhibiting a voicing bar and strong formant structure). However, there were too few (81) to analyze separately.

After examining all /t/-flap tokens, 45 were excluded from further analysis. For 17 tokens, the audio was too unclear to accurately code whether the /t/ was flapped, and 2 were excluded because of errors in the transcription. The remaining 26 were excluded because they were actually glottalized tokens, mostly occurring before syllabic nasals. There were no examples of a stop realization in the remaining tokens. The revised numbers of tokens, after these exclusions, are given in Table 5.

<INSERT Table 5 ABOUT HERE>

Some errors may still persist in the coding of /t/ and /d/ as having stop realizations versus flap and approximant realizations. However, the specific realization of /t/ and /d/ are of less importance here than the effects they have on the preceding /ay/. And as will be demonstrated in the next section, the contexts coded as “flap” and the contexts coded as “faithful” have, on aggregate, different effects on what would be the phonetic precursors for /ay/ raising. Even if these segments are entirely mischaracterized as “flaps” and “stops”, they still serve as contexts with differential phonetic conditioning.

4.2. PHONETIC PRECURSORS. There are two main contenders to be the phonetic precursors of pre-voiceless /ay/ raising. The first, proposed by Moreton and Thomas (2007), is that before voiceless consonants, the offglide of the diphthong is peripheralized. The process of /ay/ raising is thus a result of the diphthong nucleus assimilating to the peripheralized glide. The second, originally proposed by Joos (1942), is that pre-voiceless shortening does not allow enough time for the full gesture from a low-back nucleus to a high front glide, so in reaction, the nucleus raises.

Additional phonetic information beyond point estimates of F1 and F2 are necessary to see how these precursors are distributed across contexts. Specifically, full formant tracks are needed to see how the /ay/ glides are affected by flapping, as well as need vowel durations. Both of these kinds of data are available from FAVE-extract but they are unfortunately not as high quality as the F1 and F2 point estimates. First of all, FAVE-extract optimizes LPC parameters to arrive at the most likely formant point estimates, but this does not necessarily produce a high quality formant track. This is especially true for a vowel like /ay/, where the high F1 and low F2 at the nucleus would be best estimated with a larger number of poles than the low F1 and high F2 in the glide. The result is that further into the glide, F2 is often poorly tracked. As for vowel duration, the forced-alignments from FAVE-align have not been hand corrected. Alignment errors may be a problem, but Yuan and Liberman (2008) found that most errors based on the acoustic models FAVE-align uses are less than 50ms. However, the aligner only has a precision of 10ms. That is, a phone's duration can increase only by increments of 10ms, and this may pose a problem for finer grained analysis necessary to determine the phonetic precursors of /ay/ raising. However, it is possible to triangulate between the qualitative generalities of the data from the PNC and results from the literature to arrive at the most likely phonetic situation at the beginning of the change.

4.2.1. OFFGLIDE PERIPHERALIZATION. Using a slightly modified version of FAVE-extract, I extracted F1 from the relevant tokens from 10% to 90% of the duration at 5% intervals. From these full formant tracks, 80% of the vowel's duration will be taken to be indicative of the glide target. 80% was selected to be as close as possible to the glide offset while simultaneously avoiding co-articulation with the following segment. Moreton (2004) and Moreton and Thomas (2007) used a more principled approach of selecting F2 maximum as the point of the glide measurement. In this data set, that approach was too fragile with respect to formant tracking errors, frequently choosing points too early in the vowel as the glide offset. Figure 8 plots the height of /ay/ glides over speakers' dates of birth, comparing pre-/t/ and /d/ /ay/ in both flapping and faithful contexts.

<INSERT *Figure 8* ABOUT HERE>

Looking at the beginning of the 20th century, there is a clear effect of offglide peripheralization for pre-faithful-/t/ only. The height of the /ay/ glide before both variants of /d/ and before flapped /t/ all appear to have roughly the same height. There is a striking raising of the /ay/ glide before flapped /t/ across the 20th century, but this could simply be the result of the raising of the /ay/ nucleus in this context. If the degree of undershoot to the glide remained constant across the 20th century, the mere fact that the nucleus rose would have the effect of also raising the glide. In order to factor out the confounding effect that raising the vowel nucleus will also have an effect of raising the glide, I calculated the average difference between maximum F1 and F1 at 80% of the vowel's duration for every speaker in these four contexts. This gives us a measure of the glide's relative peripheralization with respect to the nucleus. Figure 9 plots the diachronic trajectories for these differences.

<INSERT *Figure 9* ABOUT HERE>

Briefly looking at the trends for /ay/ before faithful /t/ and /d/ alone, depicted by the solid lines in Figure 9, it would look like the offglide peripheralization hypothesis is strongly confirmed. At the onset of the 20th century, there is a large difference between the nucleus and glide for /ay/ before /t/, and a much smaller one before /d/. There is then a sharp, S-shaped, curve whereby the distance between nucleus and glide is reduced for /ay/ before /t/, such that the

glide's relative peripheralization comes closer in line with that of /ay/ before /d/. The relative peripheralization was reduced primarily by the raising of the /ay/ nucleus towards the glide rather than the other way around. However, the relative glide peripheralization for /ay/ before flapped /t/ is indistinguishable from the pre-/d/ context. That is, there doesn't appear to be any offglide peripheralization effect at all for /ay/ when preceding flaps, regardless of their underlying status. This is actually to be expected if the reason for offglide peripheralization is the "spread of facilitation" as Moreton (2004) suggested. Moreton (2004) construes this spread of facilitation as being a low level factor involving "[n]euromuscular coupling between temporally overlapping vowel and consonant articulations." The relationship between any particular [r] and /t/ must be at the cognitive level, relatively far removed from neuromuscular planning, so it is not surprising that there would be no offglide peripheralization before flapped /t/.

These results are concordant with what Rosenfelder (2005) found in Victoria, British Columbia. While she was not similarly focused on the interaction of raising and flapping, she did report the nucleus and glide measurements for /ay/ preceding /t/-flaps separately. Figure 10 plots the means reported in her appendices. While the glide of /ay/ before /t/-flaps is a bit higher and fronter than the glide before voiced consonants, these glides are more similar to each other than to the glide preceding voiceless consonants. This is even more surprising if it is taken into account that the nuclei of pre-/t/-flap /ay/ and pre-voiced /ay/ are extremely different.

<INSERT Figure 10 ABOUT HERE>

On the other hand, Kwong and Stevens (1999) did find /ay/ offglide peripheralization was not entirely neutralized before flapped /t/. For most of their speakers, the glide's F1 was lower and its F2 was higher before a /t/-flap than before a /d/-flap. However, they did not report formant estimates for /ay/ preceding faithful /t/ and /d/, making comparison to the data from the PNC and Rosenfelder (2005) difficult. The crux of the matter is not whether there is any offglide peripheralization before /t/-flaps, but rather whether there is enough to drive /ay/ raising before /t/-flaps on phonetic grounds.

Between the data available from the PNC and the data reported by Rosenfelder (2005), it appears as if pre-voiceless glide peripheralization is largely neutralized before flaps. If the

contexts in which /ay/ raising took place were defined on the basis of offglide peripheralization, it would be expected to see it occurring only before faithful /t/.

4.2.2. PRE-VOICELESS SHORTENING. Figure 11 plots the diachronic trends for /ay/ durations from the PNC in the four contexts in question. Looking at the early part of the 20th century, voicing effects on duration are not completely neutralized before flaps, but they are heavily mitigated towards the short end of the spectrum. Figure 11 summarizes the mean duration in milliseconds from speakers born before 1920 for /ay/ preceding /t/ and /d/ in flapping and faithful contexts. When arranging contexts from shortest to longest, it looks like when preceding a /d/-flap, the duration of /ay/ is more similar to a faithful /t/ than to a faithful /d/.

<INSERT *Figure 11* ABOUT HERE>

<INSERT Table 6 ABOUT HERE>

Table 7 displays the same mean durations by following segment and context, emphasizing the duration differences between following /t/ and /d/ within contexts. The duration difference is more than two times greater in the faithful context than in the flapping context.

<INSERT Table 7 ABOUT HERE>

Braver (2014) recently found an even greater amount of duration neutralization than has been found in the PNC. Using lab speech (and, unfortunately for comparison's sake, not /ay/), Braver found that among his 12 subjects, the largest durational difference before /d/-flaps from /t/-flaps that any of them made was about 15ms. He doesn't report any durations from non-flapped /t/ or /d/, but the vowel durations he reports for pre-/d/-flap vowels are, at most, 140ms. So it appears that in Braver's data, the direction of incomplete neutralization is also towards shorter end of the duration spectrum.

It's not immediately clear why the durational distributions in the PNC should be so different from Braver (2014). It may be due to dialectal differences between the PNC speakers and Braver's speakers, it may be an /ay/ specific effect, or it may have to do with the methodological and analytic difference between these studies. Regardless, what both sets of results find is that

vowels in general, and /ay/ in particular, are shorter before /d/-flaps than before faithful /d/, perhaps nearly (but not statistically) identical to vowels before faithful /t/. This places the predicted participation of /ay/ before /d/-flaps in the raising change in an ambiguous position if it is driven by phonetic length. Before /d/-flaps, /ay/ would either participate to the same degree as /ay/ before faithful /t/, or perhaps would participate a bit more weakly.

4.2.3. PHONETIC PRECURSORS SUMMARY. The two main contenders for phonetic precursors of /ay/ raising actually produce different predictions for how /ay/ should behave before flaps. If the precursor were offglide peripheralization, as suggested by Moreton and Thomas (2007), then on the basis of the PNC data and Rosenfelder (2005), only the context of following faithful /t/ would be expected to condition the change, since the glide does not appear to be peripheralized before flaps, nor before faithful /d/. If the precursor were phonetic duration, it would be unambiguously expected that both /ay/ before faithful /t/ and before /t/-flaps would undergo the change. Our expectations for /ay/ before /d/-flaps are a bit more ambiguous. On the basis of Braver's (2014) study, it would be expected for /ay/ raising to occur at more or less the same rate before /d/-flaps as before faithful /t/ and /t/-flaps. In the PNC data, /ay/ before /d/-flaps is a bit longer than /ay/ before faithful /t/, but still much shorter than /ay/ before faithful /d/. For the purposes of further investigation, /ay/ before /d/-flaps will be categorized as a "weak undergoer" context under the duration precursor model. The categorization of contexts into "undergoer" and "non-undergoer" on the basis of these hypothesized precursors are summarized in (5).

(5) <EXAMPLE 5 HERE>

It should be noted that placing contexts into "undergoer" and "non-undergoer" categories is not really consistent with the hypocorrection model of phonetic change. Rather, it would be expected that the rate of change across contexts would vary continuously in a way proportional to the strength of the phonetic precursor in those contexts. Based on the available precursor data from the PNC, bearing in mind the necessary caveats about its quality, a more specific, quantitative, prediction can be made about how /ay/ raising ought to interact with flapping.

Taking offglide peripheralization first, the difference between F1 at 80% of the vowel's duration and maximum F1 will be used as the quantitative measure of the peripheralization precursor. We know that /ay/ raising did occur before faithful /t/, and did not occur before faithful /d/. Taking the size of the precursor before faithful /t/ as being at 100% strength, and the

size before faithful /d/ as being at 0% strength, it is possible to calculate the relative strength in the remaining contexts, which should be proportional to the degree of participation of /ay/ raising in these contexts. The estimated participation rates based on the peripheralization precursor are given in Table 8. If /ay/ were to raise before flapped /t/ and /d/, it would be expected to do so somewhere between 10% as much as /ay/ before faithful /t/. Taking the same logic and applying it to the duration precursor (Table 9), it appears that /ay/ before /t/-flaps ought to participate in the change at a higher rate than /ay/ before faithful /t/, and /ay/ before /d/-flaps ought to participate at about 60% the rate of /ay/ before faithful /t/.

<INSERT Table 8 ABOUT HERE>

<INSERT Table 9 ABOUT HERE>

Section 4.3 is devoted to describing the outcome of a non-linear model of /ay/ raising, but the results in that section can be briefly prefigured by examining a plot of speaker means of /ay/ F1 across these contexts. In Figure 12 it can be seen that the pattern of raising before /t,d/-flaps looks very similar to the pattern before faithful /t,d/. That is, the observable pattern of raising looks neither like the coarse categorization laid out in 0, nor like the quantitative predictions in Table 8 and Table 9. Rather, it appears that /ay/ raising has always been conditioned by the underlying phonological status of the following segment, not the phonetic properties of the context.

<INSERT Figure 12 ABOUT HERE>

4.3. NON-LINEAR BAYESIAN MODELING. At this point, the qualitative impression from Figure 12 requires quantitative support from statistical modeling. However, standard linear-mixed effects models will be insufficient to address the questions at hand. For example, it might be possible that /ay/ preceding flaps behaved as one of the two precursor hypotheses predict at the beginning of the change, but then underwent reanalysis to be conditioned by underlying voicing at a later point. That is the “initiation” of the change may have been governed by phonetic factors, but the “propagation” of the change by phonological factors (Janda and Joseph 2003; Ohala 2012). Simply fitting a linear model to the data would wash out any time dependent effects like this, because it would attempt to describe the data with one slope and intercept for the entire time range, while the theoretical model suggest at least two different slopes, one before reanalysis and one after. A modeling approach which allows for a non-linear relationship between the change and date of birth will be pursued. There are a number of non-linear modeling methods to choose from, including smoothing-spline ANOVAs (Davidson 2006) and generalized additive models (Wieling, Nerbonne, and Baayen 2011). In this paper, however, I will be using a Bayesian method based on Ghizta and Gelman (2014) for a number of reasons. First, the mathematical description of the model is simpler than many other non-linear modeling techniques, even if the way the parameters are estimated is more complex. Second, random effects for speakers and words are more straightforwardly integrated into the model. Finally, it is easier to generate 95% Bayesian credible intervals for all parameters and generated quantities in the model than it is using other methods, and the credible intervals are more intuitively understandable. A full mathematical description of the model is given in Appendix **Error! Reference source not found.** This description is probably longer than the descriptions of most statistical models in the literature, but that wouldn’t be the case if the full mathematical description of, say, a smoothing-spline ANOVA, or a tensor product smooth had to be included in the full text.

The single sentence description of the model is it’s a first order autoregressive model over the rate of change. The model will produce non-linear estimates for year-over-year changes, which will be labeled as δ_j , where j is an index for the date of birth. If normalized F1 of /ay/ before faithful /t/ lowered by 0.01 between 1899 and 1900, then $\delta_{1900} = -0.01$. The δ_j of any given date of birth is constrained to be similar to δ_{j-1} , i.e. the rate of change of the previous date of birth. Exactly how similar δ_j and δ_{j-1} ought to be is a parameter of the model itself, so the

smoothness or wobbliness of the non-linear aspect of the model is optimized on the basis of the data. The expected normalized F1 of /ay/ for a given year, which will be labeled as μ_j , is calculated by summing up all of the year over year changes up to the year in question. The first observable date of birth is 1889, so the estimated F1 in 1892 (μ_{1892}) is equal to $\delta_{1889} + \delta_{1890} + \delta_{1891} + \delta_{1892}$. Separate δ_j and μ_j values were estimated for each of the four contexts ([/t/, /d/]×[faithful, flap]), so that there would be no bias in the model for any of the differences between curves to shrink towards 0. Random intercepts of speaker, word, and street were also included in the model. “Street” here is a variable coding for each graduate student research group mentioned in §2.

The parameters of the statistical model described in Appendix **Error! Reference source not found.** were estimated using the No U-Turn Sampler (NUTS) (Hoffman and Gelman 2011) as implemented in Stan (Stan Development Team 2014). It is a form of Markov Chain Monte Carlo, which takes an iterative approach to estimating the probability of parameter values given the data (a.k.a. the posterior), which is proportional the probability of the data given the parameters (the likelihood) times the probability of the parameters (the prior).⁴ The Bayesian analysis done here requires specifying some kind of prior expectation for what the values of the independent parameters are. These priors can vary in terms of the restrictiveness they impose on the estimation of the independent parameters, i.e. their “informativeness”. Priors can range the full spectrum from non-informative (that is, they impose minimal restriction on the estimation of the independent parameters), to fully informative (that is, they impose stronger restriction on the estimation). Gelman et al. (2008) propose “weakly-informative” priors for logistic regression, for example. Their proposed priors take into account that coefficients in logistic regressions tend range between -5 and 5, but allow for more extreme values when justified by the data. All of the priors for this model were either non-informative or weakly-informative, meaning the parameter estimates reported below are driven most strongly by the data provided to the model, not by prior expectations about what those estimates would be.

To ensure that the sampler settled on a stable distribution, I fit 4 chains with 4000 iterations. The first 2000 iterations were discarded as a burn-in, and the Rubin-Gelman diagnostic, (\hat{R}) was used to determine convergence. For all of the parameters reported here, the \hat{R} was sufficiently close to 1 (≤ 1.1) to consider them converged.

4.3.1. RESULTS. To begin with, Figure 13 plots the estimated scale parameters (σ s) from the model as a sanity check. These values can be compared against the maximum likelihood estimates from Figure 6, and the random effects standard deviations from Table 3. The σ labelled “within speaker variation” is estimated to be about 0.51. This is slightly smaller than the residual deviance from the mixed effects model described in Table 3, but very similar to the average within-speaker standard deviation from Figure 6. The estimated between-speakers standard deviation is 0.21, which is very similar to both the rolling estimate of the inter-speaker standard deviation from Figure 6 and the standard deviation of the by-speaker random effects from Table 3. Finally, the between words standard deviation is estimated to be 0.14, which is actually a bit smaller than the standard deviation of the by-word random intercepts from Table 3, but this should perhaps not be too surprising since there are fewer word types and fewer phonological contexts represented in this model than in the full model. These scale parameters appear to be generally reasonable when compared to other similar estimates from the data, giving us some confidence that the model, as described in Appendix **Error! Reference source not found.** was reasonably specified.

<INSERT *Figure 1* ABOUT HERE>

The first model parameter which varies across time is $\delta_{[j,k]}$, which can be understood as the year-over-year differences in F1, or the rate of change. Figure 14 plots the model estimates along with a ribbon indicating the 95% Highest Posterior Density Interval, which will serve as our credible interval for the remaining analysis of this model. The interpretation of these credible intervals is different from frequentist confidence intervals. They indicate that there is a 95% probability that the value of $\delta_{[j,k]}$ lies within the interval. Rather than a tabular representation of coefficient estimates and p-values, these graphical intervals should be understood as indicating the reliability of the effect.

<INSERT *Figure 14* ABOUT HERE>

For about the first two-thirds of the time course of the change, the estimated year-over-year differences for /ay/ before /t/ (both faithful and flapped) hovers around -0.01, although it doesn't appear to be reliably different from 0 until 1920 for /ay/ preceding /t/-flaps. This is relatively

identical to the estimated slope from the full model described in Table 1, which was -0.108 per decade, or -0.0108 per year. So far, it does not look as if there was some time period where pre-flap /ay/s were either both participating, or both not participating in /ay/ raising. Rather, /ay/ preceding /d/-flaps appears to not be undergoing any change, and /ay/ before /t/-flaps appears to be undergoing the same change as /ay/ before faithful /t/. The fact the credible interval for pre-/t/-flap /ay/ doesn't exclude 0 for about 20 years after it first does for pre-faithful-/t/ is almost certainly because there is an order of magnitude more data for /ay/ before faithful /t/.

<INSERT Figure 15 ABOUT HERE>

Figure 15 plots the actual expected F1 and 95% credible intervals for /ay/ across the four contexts. Modulo the wider credible intervals in the flapping facet, which again are almost certainly due to the sparser data for pre-flap /ay/, the profile of the change is largely identical between the two contexts: /ay/ raises before underlying /t/, and does not before underlying /d/, and flapping does not seem to perturb that change.

One thing not immediately clear from Figure 15 is that there appears to be a weak main effect where /ay/ before flaps is slightly lower than before faithful /t,d/. This effect is clearer in Figure 16, which plots the same estimates from Figure 15, but this time emphasizes the difference between faithful and flapping contexts. It's not immediately clear why /ay/ preceding flaps should be slightly lower than preceding faithful /t,d/, but two things should be noted. First, it appears to affect /ay/ before /d/-flaps and /t/-flaps to a similar degree. Second, neither of the phonetic precursors considered above would predict an effect like this. If anything, the duration precursor would predict that /ay/ should be higher before flaps than before faithful realizations, and the peripheralization precursor would predict that /ay/ before /d/-flaps would be higher and /ay/ before /t/-flaps would be lower than before faithful realizations.

<INSERT Figure 16 ABOUT HERE>

Something else that is not immediately clear from a visual inspection of Figure 15 is whether the way /ay/ differentiates itself between pre-/t/ and pre-/d/ contexts is the same between faithful and flapping realizations. It could be the case that /ay/ before /t/-flaps does differentiate from /ay/

before /d/-flaps, but at a slower rate, or at a later time than it does before faithful-/t/ and faithful-/d/. The height difference between pre-/t/ and pre-/d/ contexts was not an actual parameter of the model itself, but it is possible to generate estimated differences and 95% credible intervals for those differences from the model parameters, and Figure 17 plots these estimates. It appears that the trend over time for /ay/ differentiation is nearly identical between flapping and faithful context. In fact, the 95% credible interval for the height difference begins to exclude 0 at the same time for both contexts (approximately 1915).

<INSERT Figure 17 ABOUT HERE>

It is worth re-iterating at this point that the curves for each of the four contexts investigated here were estimated separately in the model. That is, the model did not assume that there should be any similarity between any the diachronic curves plotted here. The fact that /ay/ height diverges identically in faithful and flapping contexts is a property of the data, not modeling assumptions.

4.3.2. NON-LINEAR MODELING SUMMARY. The result in Figure 17 is perhaps the most crucial one in this section. Its takeaway point is twofold. First, as soon as there was a detectable difference in height for /ay/ in pre-faithful-/t/ position and pre-faithful-/d/ position, there was also a detectable difference of the same magnitude in pre-/t/-flap and pre-/d/-flap contexts. Secondly, the overall way in which /ay/ differentiated in height across the 20th century before faithful-/t/ and faithful-/d/ appears to be the same before /t/-flaps and /d/-flaps.

4.4. ANALOGY INVESTIGATION. The primary thesis of this paper is that pre-voiceless /ay/ raising has always been conditioned by the phonological properties of its context, not the phonetic properties. However, there is one confound to the results from §4.3 that may call into question whether the effect demonstrated there was truly phonological. Almost all of words in which /ay/ appears before /t/-flaps are morphologically complex (e.g. *fighting*, *united*, *writing*). For very few roots (36 tokens in total) does /ay/ appear exclusively before /t/-flaps, and they are forms of *title*, *vitamin*, and the suffix *-itis*. It could be that /ay/ raising began in the phonetically predicted context (before faithful /t/), and then the vowel quality analogized to derived and inflected forms of the root without ever making reference to other phonological properties of the root. Such a process doesn't seem very likely in view of the results from §4.3.1, especially looking at Figure 17. The analogy would have had to be nearly instantaneous for all roots involved. Moreover, while this analogy would be very different from phonology as it is traditionally understood, it is still quite a few steps removed from the continuous properties of speech upon which the hypocorrection model is based.

However, in an effort to appropriately address this open question, I coded every word root in the flapping data for how frequently it occurs in flapping and faithful contexts according to the word frequency norms from SUBTLEX_{US}. Table 10 illustrates what this looked like for the root *unite*. For each individual word which contained the root *unite*, its frequency per 1,000 words was collected, and it was coded for whether or not the /t/ would be flapped. Then, the frequencies for each context were summed, and a ratio of flapping to faithful realizations was calculated (see Table 11). For the root *unite*, it appeared in flapping contexts about 12 times more often than in faithful contexts.

<INSERT Table 10 ABOUT HERE>

<INSERT Table 11 ABOUT HERE>

The hypothesis being pursued here is that the more frequently a root appears with /ay/ in a faithful context, where the raising change is phonetically natural, compared to a flapping context, the more likely the vowel quality is to analogize to other realizations of the root. Of course this is a relatively simplistic approach to quantifying the likelihood of analogy, but if analogy is playing a powerful enough role in producing the appearance of a phonologically conditioned phonetic change, then even an imperfect measure ought to demonstrate this effect.

As a first pass at the question, Figure 18 plots F1 means for /ay/ in flapping contexts. The data was split into two categories: roots which occur more often in flapping contexts, and roots which occur more often in faithful contexts. The impression from Figure 18 is that if there is an effect of the flapping to faithful ratio, it is a weak one.

<INSERT Table 12 ABOUT HERE>

The ratio of flapping-to-faithful frequency is a continuous factor, so it was entered into a linear mixed-effects model using a \log_2 transform.⁵ Table 12 displays the fixed-effects estimates from the model, along with 95% confidence intervals based on 10,000 semiparametric bootstrap replicates obtained using `bootMer`. The only parameters where the confidence interval excludes 0 are the Intercept, the main effect of a following /t/, and the interaction of /t/ with date of birth. The confidence interval for three way interaction of Decade×Ratio×[-voice] doesn't exclude 0, but just barely. Assuming, briefly, that the effect is actually non-zero, the size of the effect can be visualized by calculating the fitted values from each of the bootstrap replicates, and plotting from the 2.5th percentile to the 97.5th percentile, illustrating a 95% range of the fitted values. The result is Figure 19, which is difficult to interpret because the effect size is so small.

<INSERT Figure 19 ABOUT HERE>

A model comparison route to evaluating the effect of the flapping:faithful ratio suggests that there is not even a main effect. Table 13 displays the results of likelihood ratio tests, as well as the AIC and BIC for models which differ only in the way the ratio was added to the model. None

of the likelihood ratio tests are significant, and the model with the smallest AIC and BIC does not include the ratio as a predictor.

<INSERT Table 13 ABOUT HERE>

While the ratio of the frequency with which a root appeared in flapping contexts to its frequency in faithful contexts may not be the most sophisticated operationalization for analogy, any more sophisticated approach will face the challenge of successfully analogizing the vowel quality for /ay/ raising while simultaneously failing to analogize the vowel quality for /ey/ raising, which is another phonetic change that occurred in Philadelphia in the 20th century. Following consonants conditioned /ey/ raising, but this conditioning applied completely transparently. Figure 20a is an illustrative example, plotting the mean F1 for just the lexical items *day* and *days*. *Days* undergoes a phonetic change which is just slightly smaller in magnitude than pre-voiceless /ay/ raising, and no analogical force *day* to raise. For the sake of an accurate comparison, Figure 20b plots a similar illustrative plot for the lexical items *fight* and *fighting*. Both of these lexical items appear to undergo an identical raising change.

<INSERT Figure 20 ABOUT HERE>

5. DISCUSSION. Most of the analysis in this paper has, so far, been devoted to being certain about what didn't happen to /ay/. Pre-voiceless raising didn't first begin in the contexts with the strongest phonetic precursors, then subsequently generalize or analogize along phonological or lexical dimensions. In different terms, the evidence suggests that the set of environments where this change was initiated were defined on phonological grounds, rather than in terms of the phonetic properties of those environments.

The appropriate next step forward when faced with this result depends greatly on one's theoretical commitments. If our commitment to the phonetic precursors model of phonetic change extended beyond apparent counter-evidence, then the next appropriate step would be to search for new phonetic precursors that could accurately predict how this change was circumscribed. This route could prove to be long and fruitful, as the set of possible precursors is large and possibly non-finite. However, the same argument could be levied against the set of

possible phonological explanations for phonetic changes. This underlines a central problem in comparing phonetic and phonological theories for how this change occurred, as most explanations grounded in either categorical phonology or continuous phonetics will be post-hoc and highly flexible. A deductive approach to settling the question is therefore simply not open to us on the basis of current theory since the premises are not fixed. However, just because there may be infinite explanations on phonetic and phonological grounds does not mean that all explanations are equally probable. If pre-voiceless shortening and offglide peripheralization are taken to be the most likely phonetic precursors simply because they have been proposed in the literature, then the results presented here are highly unlikely on the basis of the most likely phonetic predictions. On the other hand, these results are exactly what would be expected if /ay/ raising has always been conditioned by the underlying phonological status of the following segment.

So what did happen to /ay/ in Philadelphia? These results support part of a larger argument I would like to make that in order for two contextual variants of a speech sound to diverge in their phonetics over time, they must, all else being equal, be treated as being qualitatively different categories by speakers from the moment they begin to diverge. That is, a categorical split of /ay/ into two new allophones or phonemes is not the reanalysis of a longer term phonetic change. Rather, the longer term phonetic change is only possible because /ay/ split into two new allophones or phonemes either previous to or concurrent with the onset of the phonetic change. The split allowed for their phonetic targets to be learned separately, and to change independently.

I propose that very early in the change, there were two categorically distinct, but phonetically similar variants of /ay/, distributed according to the voicing specification of the following segment, and that one of them underwent a phonetically gradual change in height, and the other remained low. These allophones differed phonetically in terms of their offglides and duration before faithful /t/ and /d/, but when those phonetic differences were neutralized in the flapping context, it was still only the pre-voiceless one which underwent the change. This is a more extreme version of the Big Bang theory of sound change put forward by Janda and Joseph (2003). They propose that purely phonetic factors guide sound changes very briefly, and are eventually overridden by phonological conditioning. There is, in fact, no detectable period where the pattern of /ay/ raising aligned with what would be predicted on purely phonetic grounds. The conclusion I draw is that either the period of purely phonetic conditioning was too brief to be

identified, or was non-existent. If the situation is the former, then that means there is a greater challenge than perhaps has been appreciated in identifying phonologization *in vivo*. If the situation is the latter, then the question arises as to how these two categorically different variants came to exist.

Under my proposal, there must have been a categorical difference between pre-voiced and pre-voiceless /ay/ at the onset of the change, but the nature of that difference is, unfortunately, not specified by my model. We can't differentiate between proposals that place the distinction in the underlying representation (Mielke, Armstrong, and Hume 2003) and those that generate it in the phonological grammar (Idsardi 2006) or those that might mix the two (Pater 2014). However, Bermúdez-Otero (2013) makes a compelling argument that a categorical distinction between pre-voiced and pre-voiceless /ay/ is not necessarily a new development. Rather, he argues that a long-standing categorical process of pre-fortis clipping is responsible for producing two allophones of /ay/. As was already demonstrated, the selection of contexts to undergo /ay/ raising is not proportional to phonetic duration, but pre-fortis clipping is construed here as a categorical, phonological process, even if its primary phonetic consequence is a shorter vowel duration. Under Bermúdez-Otero's account, clipped /ay/, [ǎy], was usually shorter than full /ay/, [āy]. But under certain circumstances (like when preceding flaps), this phonological distinction is phonetically neutralized to have the same, short, duration. The analysis in Section 4 strongly indicates that it was the phonological status of [ǎy] that selected it for raising, rather than its phonetic properties in any given context.

Whether phonetic /ay/ raising necessitates a reorganization in the phonological grammar of Philadelphians depends on how much or how little one wants to make their phonology dictate phonetics. For example, there may have only been one phonological process in the grammar across the entire 20th century, which could be given as (6).

(6) CLIPPING ay → ǎy/___ -voice

The substantive change observed in this paper would thus be a shift in the phonetic realization of [ǎy], which used to just be realized with a shorter duration, but then began to also exhibit a change in its height. Alternatively, a context-free process could have been introduced to the grammar which altered the phonological specification of height for [ǎy], which could be given as (7).

(7) CLIPPING ay → äy/___ -voice

RAISING äy → -low

This change in the phonological specification of [äy] resulted in the observed gradual phonetic shift. An additional possibility is that these two phonological grammars, in addition to any others which might result in similar phonetic outcomes, were all covertly being used in a mixture, such as Mielke, Baker, and Archangeli (forthcoming) have found for the distribution of bunched and retroflex /r/ (which are largely acoustically indistinguishable) in American English. Additional diagnostics to differentiate between these possibilities, like other phonological processes which interact with vowel height, are not forthcoming, so at the moment this discussion will have to be set aside.

Regardless of how different phonological categories are being represented or generated, my conclusion that they are not products of the phonetic change, but rather necessary ingredients for the change to happen, reopens most of the questions that hypocorrection was supposed to resolve. Gradual assimilation of the vowel's nucleus to a peripheralized glide, for example, would explain why the change happened at all. Without this explanation, the mystery of why [äy] and [äy] didn't just both maintain their low nuclei forever remains. Perhaps a maximal dispersion theory could be turned to salvage the situation. For example, Boersma and Hamann (2008) explicitly model maximal acoustic dispersion as being a product of cross generational change. However, we are again left with the same, fundamentally difficult Actuation Problem: Why now? Why never before? Why here? Why not everywhere? Since a definitive answer to the Actuation Problem has not been provided in the nearly 50 years since it was first given a name, I can hopefully be forgiven for not settling the issue here.

As for the larger argument I want to make regarding the early influence of phonology on phonetic changes, a definitive case for it will require more examples of phonetic changes from more dialects investigated in similar depth as I have done here. I have focused exclusively on /ay/ raising in order to provide it the thorough treatment necessary to establish with relative certainty that for this change, phonological conditioning was present at its outset. Obviously generalizing too broadly from one specific case should be avoided, but I hope to have at least laid the groundwork for establishing a direction of inquiry, and sufficiently problematized the widely held conventional wisdom regarding changes of this sort.

6. CONCLUSION. The hypocorrection model of conditioned phonetic change relies crucially on the presence of phonetic precursors that drive the change. In the case of pre-voiceless /ay/ raising, I established how the two phonetic precursors which have been proposed for it (pre-voiceless shortening, and offglide peripheralization) predict /ay/ raising ought to interact with /t, d/ flapping. I found that in 20th century Philadelphia, neither set of predictions are borne out. Rather, it appears that /ay/ raising has always been conditioned by the phonological voicing of the following segment, not the phonetic properties of the context. I argued that this calls into question the hypocorrective model of phonetic change as a universal explanation, and that the possibility of early involvement of phonology ought to be explored in more sound changes.

REFERENCES:

- ADANK, PATTI; ROEL SMITS; and ROELAND VAN HOUT. 2004. A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America* 116.3099-3177. doi:10.1121/1.1795335.
- BAKER, ADAM; DIANA ARCHANGELI; and JEFF MIELKE. 2011. Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change* 23.347-74. doi:10.1017/S0954394511000135.
- BARR, DALE J.; ROGER LEVY; CHRISTOPH SCHEEPERS; and HARRY J. TILY. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68.255-78. doi:10.1016/j.jml.2012.11.001.
- BATES, DOUGLAS; MARTIN MAECHLER; BEN BOLKER; AND STEVEN WALKER. 2014. lme4: Linear mixed-effects models using Eigen and S4. <http://cran.r-project.org/package=lme4>.
- BERMÚDEZ-OTERO, RICARDO. 2007. Diachronic phonology. In *The Cambridge Handbook of Phonology*, edited by Paul de Lacy, 497-517. Cambridge: Cambridge University Press.
- BERMÚDEZ-OTERO, RICARDO. 2013. Philadelphia /aɪ/-raising without rule insertion. Paper Presented at the Symposium on Historical Phonology.
- BLEVINS, JULIETTE. 2004. *Evolutionary phonology: The emergence of sound patterns*. New York: Cambridge University Press.
- BLEVINS, JULIETTE. 2008. Natural and unnatural sound patterns: A pocket field guide. In *Naturalness and Iconicity in Language*, edited by Klaas Willems and Ludovic De Cuypere. 121-48. John Benjamins.
- BOERSMA, PAUL, and SILKE HAMANN. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25.217-70.
- BRAVER, AARON. 2014. Imperceptible Incomplete Neutralization : Production , Non-Identifiability, and Non-Discriminability in American English Flapping. *Lingua*. 152.24-44. doi:10.1016/j.lingua.2014.09.004
- BRYLSBAERT, MARC, and BORIS NEW. 2009. Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41.977-90. doi:10.3758/BRM.41.4.977.
- CHAMBERS, J. K. 1973. Canadian Raising. *Canadian Journal of Linguistics* 18.131-35.

- CONN, JEFFERY. 2005. Of ‘moice’ and men: The evolution of a male-led sound change. University of Pennsylvania dissertation.
- DAILEY-O’CAIN, JENNIFER. 1997. Canadian raising in a midwestern U.S. city. *Language Variation and Change* 9.107-20. doi:10.1017/S0954394500001812
- DARLINGTON, RICHARD B. 1970. Is Kurtosis Really ‘Peakedness?’ *The American Statistician* 24.19-22. doi: 10.2307/2681925
- DAVIDSON, LISA. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America* 120.407. doi:10.1121/1.2205133.
- EVANINI, KEELAN. 2009. The permeability of dialect boundaries: A case study of the region surrounding Erie, Pennsylvania. University of Pennsylvania dissertation.
- FRUEHWALD, JOSEF. (2008). The spread of raising: Opacity, lexicalization, and diffusion. *Penn Working Papers in Linguistics* 14.83-92.
- FRUEHWALD, JOSEF. 2013. The Phonological Influence on Phonetic Change. University of Pennsylvania dissertation.
- GARRETT, ANDREW, and KEITH JOHNSON. 2013. Phonetic bias in sound change. In Yu, 51-97.
- GELMAN, ANDREW; ALEKS JAKULIN; MARIA GRAZIA PITTAU; and YU SUNG SU. 2008. A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics* 2 (4): 1360-83. doi:10.1214/08-AOAS191.
- GHITZA, YAIR, and ANDREW GELMAN. 2014. The Great Society, Reagan’s Revolution, and Generations of Presidential Voting. ms.
http://graphics8.nytimes.com/newsgraphics/2014/07/06/generations2/assets/cohort_voting_20140707.pdf.
- GORDON, ELIZABETH; MARGARET MACLAGAN; and JENNIFER HAY. 2007. The ONZE Corpus. In *Models and methods in the handling of unconventional digital corpora. Volume 2: Diachronic corpora*. 82-104. Palgrave Macmillan.
- HARRINGTON, JOHN; SALLYANNE PALETHORPE; and CATHERINE I WATSON. 2000. Does the Queen speak the Queen’s English? *Nature* 408.927-928. doi:10.1038/35050160.
- HARRINGTON, JONATHAN; FELICITAS KLEBER, and ULRICH REUBOLD. 2007. /u/-fronting in RP: A link between sound change and diminished perceptual compensation for coarticulation. *ICPhS XVI*. 1473-76.

- HAY, JENNIFER B.; JANET B. PIERREHUMBERT; ABBY J. WALKER; and PATRICK LASHHELL. 2015. Tracking word frequency effects through 130 years of sound change. *Cognition* 139.83-91. doi:10.1016/j.cognition.2015.02.012.
- HAYES, BRUCE, and DONCA STERIADE. 2004. Introduction: The phonetic bases of phonological Markedness. In *Phonetically Based Phonology*, edited by Bruce Hayes, Robert Kirchner, and Donca Steriade. 1-33. Cambridge University Press. doi:10.1017/cbo9780511486401.001
- HOFFMAN, MATTHEW D, and ANDREW GELMAN. 2011. The No-U-Turn Sampler: Adaptively setting path lengths in Mamiltonian Monte Carlo." *ArXiv.org*. <http://arxiv.org/abs/1111.4246>.
- IDSARDI, WILLIAM. 2006. "Canadian Raising, Opacity, and Rephonemicization." *The Canadian Journal of Linguistics*. 51.119-26. doi:10.1353/cjl.2008.0007.
- JANDA, RICHARD D, and BRIAN D JOSEPH. 2003. Reconsidering the canons of sound-change: Towards a 'big bang' theory." In *Historical Linguistics 2001. Selected papers from the 15th International Conference on Historical Linguistics, Melbourne*, edited by Barry Blake and Kate Burridge, 205-19. John Benjamins.
- JOHNSON, DANIEL EZRA. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistic Compass* 3.359-83. doi:10.1111/j.1749-818X.2008.00108.x
- JOOS, MARTIN. 1942. A phonological dilemma in Canadian English. *Language* 18.141-44. doi: 10.2307/408979
- KRUSCHKE, JOHN K. 2011. *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Oxford: Elsevier.
- KWONG, KATHERINE, and KENNETH STEVENS. 1999. On the voiced-voiceless distinction for writer/rider. *Speech Communication Group Working Papers (Research Laboratory of Electronics at MIT)* 11.1-20.
- LABOV, WILLIAM. 1963. The social motivation of a sound change. *Word* 19.273-309.
- LABOV, WILLIAM. 2001. *Principles of linguistic change. Volume 2: Social Factors. Language in Society*. Oxford: Blackwell.
- LABOV, WILLIAM. 2007. Transmission and Diffusion. *Language* 83.344-87. doi:10.1353/lan.2007.0082.
- LABOV, WILLIAM, and INGRID ROSENFELDER. 2011. The Philadelphia Neighborhood Corpus.

- LABOV, WILLIAM; SHERRY ASH; and CHARLES BOBERG. 2006. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. New York: Mouton de Gruyter.
- LABOV, WILLIAM; INGRID ROSENFELDER; and JOSEF FRUEHWALD. 2013. One hundred years of sound change in Philadelphia: Linear incrementaion, reversal, and reanalysis. *Language* 89.30-65.
- MIELKE, JEFF; MIKE ARMSTRONG; and ELIZABETH HUME. 2003. Looking through opacity. *Theoretical Linguistics* 29.123-39.
- MIELKE, JEFF; ADAM BAKER; and DIANA ARCHANGELI. 2015. Individual-level contact limits phonological complexity: evidence from bunched and retroflex /ɹ/. *Language*, to appear.
- MORETON, ELLIOTT, and ERIK R THOMAS. 2007. Origins of Canadian Raising in voiceless-coda effects: A case study in phonologization. In *Papers in Laboratory Phonology 9*, edited by Jennifer Cole and Jose Ignacio Hualde, 37-64. Cambridge University Press.
- MORETON, ELLIOTT. 2004. Realization of the English postvocalic [voice] contrast in F1 and F2. *Journal of Phonetics* 32.1-33. doi:10.1016/S0095-4470(03)00004-4.
- OHALA, JOHN J. 1981. The listener as a source of language change. In *Papers from the Parasession on Language and Behavior*, edited by C S Masek, R A Hendrick, and M F Miller. Chicago Linguistics Society.
- OHALA, JOHN J. 1990. The phonetics and phonology of aspects of assimilation. In *Papers in Laboratory Phonology 1: Between the Grammar and Physics of Speech*, edited by John Kingston and Mary E Beckman, 258-82. Cambridge: Cambridge University Press.
- OHALA, JOHN J. 2012. The listener as a source of sound change: An update. In *The Initiation of Sound Change: Perception, Production, and Social Factors*, edited by Maria-Josep Solé and Daniel Recasens, 21-35. Philadelphia: John Benjamins.
- OHALA, JOHN J., and MANJARI OHALA. 1991. Reply to Commentators. *Phonetica* 48.271-74. doi:10.1159/000261892.
- PATER, JOE. 2014. Canadian raising with language-specific weighted constraints. *Language* 90.230-40. doi:10.1353/lan.2014.0009.
- PIERREHUMBERT, JANET B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In *Frequency and the Emergence of Linguistic Structure*, edited by Joan Bybee and Paul Hopper, 45:137-57. Typological Studies in Language. John Benjamins.

- PITT, MARK A; LAURA DILLEY; KEITH JOHNSON; SCOTT KIESLING; WILLIAM RAYMOND; ELIZABETH HUME; and ERIC FOSSLER-LUSSIER. 2007. Buckeye Corpus of Conversational Speech (2nd release). Columbus, OH: Department of Psychology, Ohio State University. www.buckeyecorpus.osu.edu.
- RATHCKE, TAMARA, and JANE STUART-SMITH. 2014. On the impact of noise on vowel formant measures. Paper presented at Methods in Dialectology XV.
- ROSENFELDER, INGRID. 2005. Canadian Raising in Victoria , B. C.: an Acoustic Analysis. Ruprecht Karl University of Heidelberg dissertation.
- ROSENFELDER, INGRID, JOSEF FRUEHWALD, KEELAN EVANINI, SCOTT SEYFARTH, KYLE GORMAN, HILARY PRICHARD, and JIAHONG YUAN. 2014. FAVE (Forced Alignment and Vowel Extraction) v1.1.3. doi:10.5281/zenodo.9846.
- SANKOFF, GILLIAN. 2006. Age: Apparent time and real time. In *Encyclopedia of Language and Linguistics*. Oxford: Elsevier.
- SANKOFF, GILLIAN, AND HÉLÈNE BLONDEAU. 2007. Language change across the lifespan: /r/ in Montral French. *Language* 83.560-88.
- SMITH, JENNIFER; MERCEDES DURHAM; and LIANE FORTUNE. 2007. ‘Mam, my trousers is fa’ in doon!’: Community, caregiver, and child in the acquisition of variation in a Scottish dialect. *Language Variation and Change* 19.63-99. doi: 10.1017/s0954394507070044
- STAN DEVELOPMENT TEAM. 2014. Stan: A C++ library for probability and sampling. <http://mc-stan.org/>.
- TUCKER, R WHITNEY. 1944. Notes on the Philadelphia dialect. *American Speech* 19.37-42. doi:10.2307/486530
- VANCE, TIMOTHY. 1987. ‘Canadian Raising’ in some dialects of the northern United States. *American Speech* 62.195-210. doi:10.2307/454805
- WAGNER, SUZANNE E. 2007. We act like girls and we don’t act like men: The use of the male-associated variable (ay0) in South Philadelphia. *Penn Working Papers in Linguistics*, 13.1: 393-406.
- WARNER, NATASHA, and BENJAMIN V. TUCKER. 2011. Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America* 130.1606-1617. doi:10.1121/1.3621306.

- WEDEL, ANDREW B. 2006. Exemplar models, evolution and language change. *Linguistic Review* 23.247-74. doi:10.1515/TLR.2006.010.
- WEDEL, ANDREW B. 2007. Feedback and regularity in the lexicon. *Phonology* 24.147-185. doi:10.1017/S0952675707001.
- WEINREICH, URIEL; WILLIAM LABOV; and MARVIN HERZOG. 1968. Empirical foundations for a theory of language change. In *Directions for Historical Linguistics*, edited by W Lehmann and Y Malkiel. U. of Texas Press.
- WIELING, MARTIJN; JOHN NERBONNE; and R HARALD BAAYEN. 2011. Quantitative social dialectology: explaining linguistic variation geographically and socially. *PloS One* 6 (9): e23613. doi:10.1371/journal.pone.0023613.
- WOOD, SIMON N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* 73.3-36. doi:10.1111/j.1467-9868.2010.00749.x/full.
- WOOD, SIMON N. 2014. *mgcv*. <http://cran.r-project.org/web/packages/mgcv/>.
- YU, ALAN C. L. 2013a. Individual differences in soci-cognitive processing and the actuation of sound change. In Yu, 201-27.
- YU, ALAN C. L. (ed). 2013b. *Origins of sound change: Approaches to phonologization*. Oxford University Press.
- YUAN, JIAHONG; and MARK LIBERMAN. 2008. Speaker identification on the SCOTUS corpus. *The Journal of the Acoustical Society of America*, 123.3878-3882. doi:10.1121/1.2935783
- ZELLOU, GEORGIA; and MEREDITH TAMMINGA. 2014. Nasal coarticulation changes over time in Philadelphia English. *Journal of Phonetics* 47.18-35. doi:10.1016/j.wocn.2014.09.002.

NOTES

1. Here and throughout, when a figure plots mean formant values for a speaker, they were calculated by first taking the mean formant value for each word for each speaker, and then calculating the mean category value on the basis of these by-word means. This is an attempt to reduce any undue effect of highly frequent words, like and right on the estimation of the mean. Using this method, highly frequent words and only get to contribute once to the estimation of speakers' means, as do low frequency words.
2. A model with maximal random effects (Barr et al. 2013) would include random slopes for the interaction and main effects of frequency by speaker: (following_voicing*log2freq|Speaker). This model proved computationally intractable to fit. The omission of these random slopes may have an anti-conservative effect on the estimation of the main effect of frequency and its interaction with voicing context (D. E. Johnson 2009). That is, the true size of these effects may be smaller than the model has estimated. Since the interaction of frequency with the voicing context and date of birth is already a null result, the successful inclusion of these random slopes would not likely alter the outcome qualitatively.
3. I compared the effect of doubling frequency to the rate of change of pre-voiceless /ay/, which is approximately equal to the effect labeled "Decade × [-voice]" = -0.106. A word with double the frequency of another word will have a normalized F1 approximately -0.013 lower, which is about $\frac{1}{10}$ th the effect of a decade in date of birth, or more simply 1 year.
4. See Kruschke (2011) for an accessible introduction to MCMC and Bayesian Modeling.
5. Using the log₂ transform means that a flap:faithful ratio of 2:1 has a value of 1, a 1:1 ratio has a value of 0, and a 1:2 ratio has a value of -1.

FIGURES

Figure 1 Hypothetical rates of /u/ fronting under different phonetic bias conditions.....	49
Figure 2 The year of interview and subject's age in the white, Philadelphia born subset of the PNC. A slight jitter (uniform noise between [-0.4, 0.4]) has been added to the year of study.	50
Figure 3 Raising of pre-voiceless /ay/. There are two points for each speaker, representing their mean normalized F1 of /ay/ in the two contexts. The smoothing lines are penalized cubic regression splines. Horizontal dotted lines representing average normalized F1 for [] and [] are included for scale.	51
Figure 4 The effect of following voicing (Cohen's d), calculated for every speaker. The smoothing lines are penalized cubic regression splines.....	52
Figure 5 Estimated kurtosis of individual speakers' pre-voiceless /ay/ distributions along F1. The smoothing lines are penalized cubic regression splines. The horizontal grey line at $y=3$ represents the kurtosis of a normal distribution. The y-scale is logarithmic.	53
Figure 6 Comparison of within-speaker variation (one point per speaker) to between-speaker variation (grey line). The smoothing lines are penalized cubic regression splines. Between-speaker standard deviation calculated based on a moving 20 year date of birth window at 5 year increments. The y-scale is logarithmic.	54
Figure 7 Predicted F1 of pre-voiceless /ay/ for different Date of Birth cohorts at different ages. $\text{gam}(\text{NormalizedF1} \sim \text{ti}(\text{age}) + \text{ti}(\text{dob}) + \text{ti}(\text{age}, \text{dob}))$	55
Figure 8 /ay/ glide height across the 20th century. The smoothing lines are penalized cubic regression splines. Dark grey lines correspond to /d/; light grey lines correspond to /t/. Solid lines correspond to faithful (non-flapping) realizations of /t, d/; dashed lines correspond to flapped realizations of /t, d/.	56
Figure 9 Distance from /ay/ nucleus (maximum F1) to glide (80% of vowel duration) across the 20th century. The smoothing lines are penalized cubic regression splines. Dark grey lines correspond to /d/; light grey lines correspond to /t/. Solid lines correspond to faithful (non-flapping) realizations of /t, d/; dashed lines correspond to flapped realizations of /t, d/.	57
Figure 10 Nucleus and glide measurements of pre-voiced, pre-voiceless, and pre-/t/-flap /ay/ from Victoria, B.C. Data from Rosenfelder (2005).	58

Figure 11 Diachronic trajectory for /ay/ duration across the 20th century. The smoothing lines are penalized cubic regression splines. Dark grey lines correspond to /d/; light grey lines correspond to /t/. Solid lines correspond to faithful (non-flapping) realizations of /t, d/; dashed lines correspond to flapped realizations of /t, d/.....	59
Figure 12 Speaker F1 means for /ay/ preceding different /t/ and /d/ realizations.....	60
Figure 13 Estimated standard deviations from the Stan model. Intervals represent the range of the 95% Highest Posterior Density Interval.	61
Figure 14 Model estimates and 95% credible intervals for the year-over-year differences (δ) i.e. the rate of change. Dark grey lines correspond to /d/; light grey lines correspond to /t/. Solid lines correspond to faithful (non-flapping) realizations of /t, d/; dashed lines correspond to flapped realizations of /t, d/.....	62
Figure 15 Model estimates and 95% credible intervals for normalized /ay/ F1. Dark grey lines correspond to /d/; light grey lines correspond to /t/. Solid lines correspond to faithful (non-flapping) realizations of /t, d/; dashed lines correspond to flapped realizations of /t, d/.....	63
Figure 16 Model estimates and 95% credible intervals for normalized /ay/ F1. Dark grey lines correspond to /d/; light grey lines correspond to /t/. Solid lines correspond to faithful (non-flapping) realizations of /t, d/; dashed lines correspond to flapped realizations of /t, d/.....	64
Figure 17 Estimated /ay/ height difference between pre-/t/ and pre-/d/ contexts. Solid lines correspond to faithful (non-flapping) realizations of /t, d/; dashed lines correspond to flapped realizations of /t, d/.....	65
Figure 18 Estimated F1 of /ay/ in flapping contexts, divided by roots which appear more often in flapping contexts, and those which appear more often in faithful contexts. The smoothing lines are penalized cubic regression splines.....	66
Figure 19 95% percentiles of the fitted values from 10,000 bootstrap replicates.....	67
Figure 20 Mean F1 over date of birth for different lexical items. The smoothing lines are penalized cubic regression splines.....	68

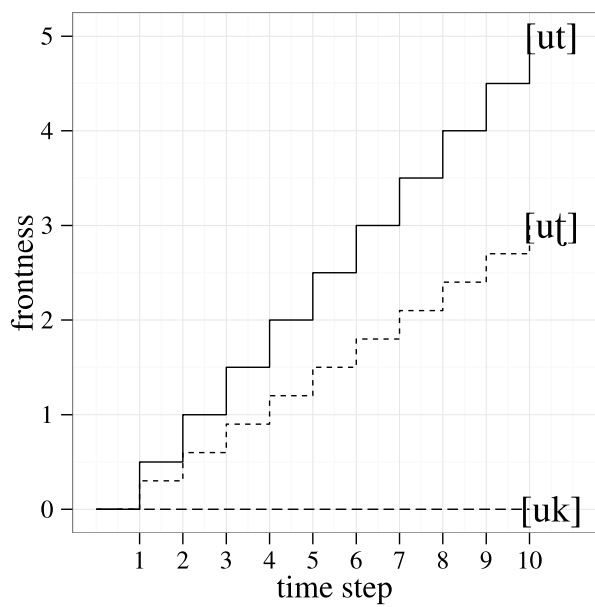


Figure 1 Hypothetical rates of /u/ fronting under different phonetic bias conditions.

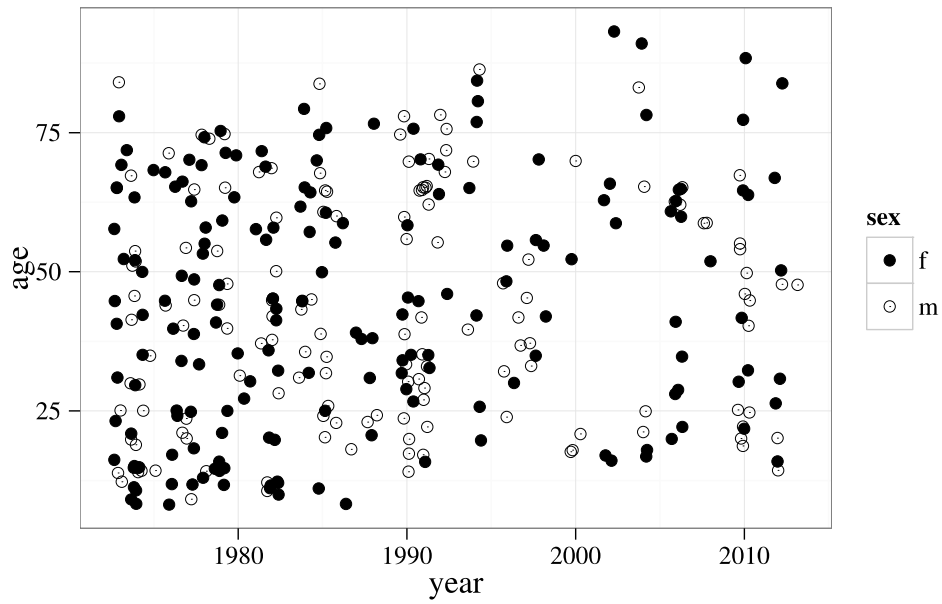


Figure 2 The year of interview and subject's age in the white, Philadelphia born subset of the PNC. A slight jitter (uniform noise between $[-0.4, 0.4]$) has been added to the year of study.

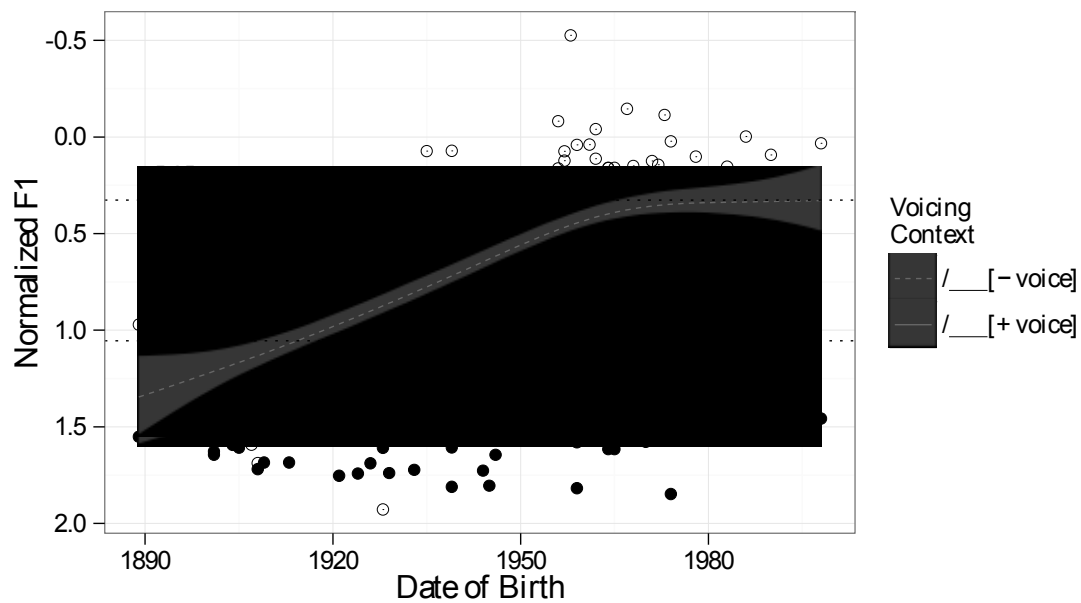


Figure 3 Raising of pre-voiceless /ay/. There are two points for each speaker, representing their mean normalized F1 of /ay/ in the two contexts. The smoothing lines are penalized cubic regression splines. Horizontal dotted lines representing average normalized F1 for [ɑ] and [ʌ] are included for scale.

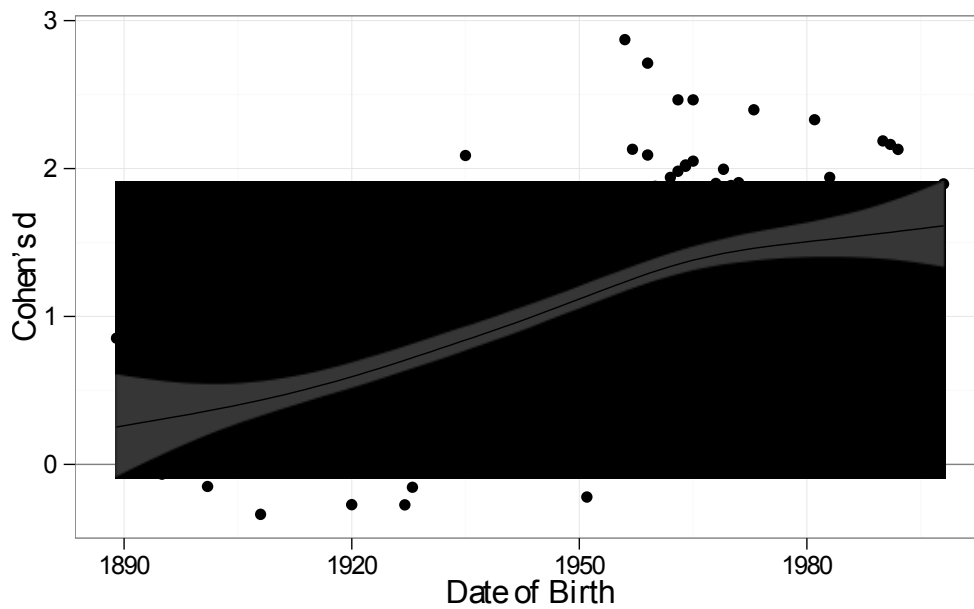


Figure 4 The effect of following voicing (Cohen's d), calculated for every speaker. The smoothing lines are penalized cubic regression splines.

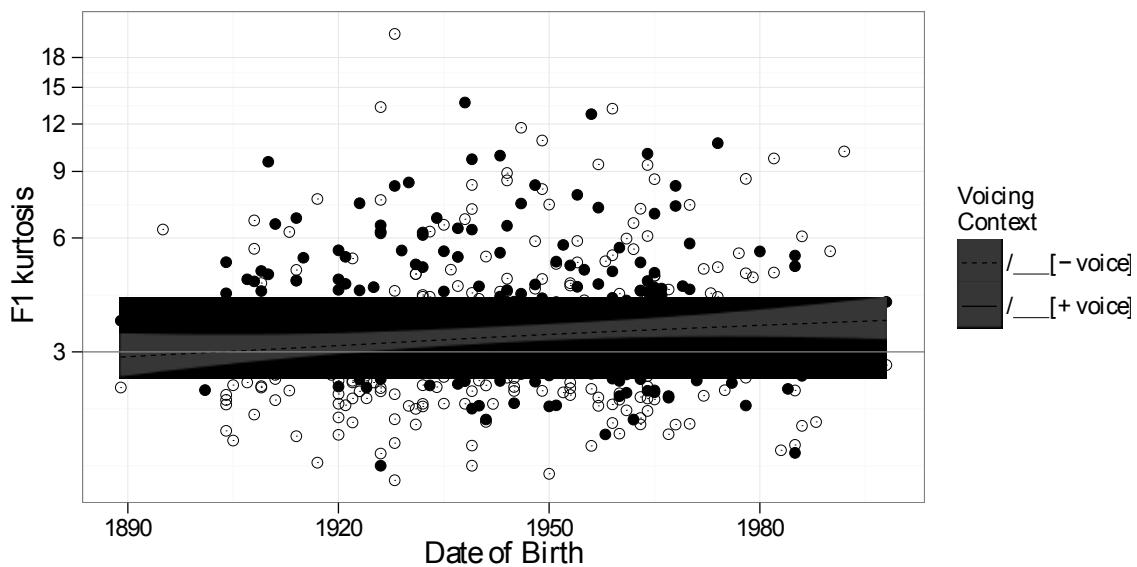


Figure 5 Estimated kurtosis of individual speakers' pre-voiceless /ay/ distributions along F1. The smoothing lines are penalized cubic regression splines. The horizontal grey line at $y=3$ represents the kurtosis of a normal distribution. The y-scale is logarithmic.

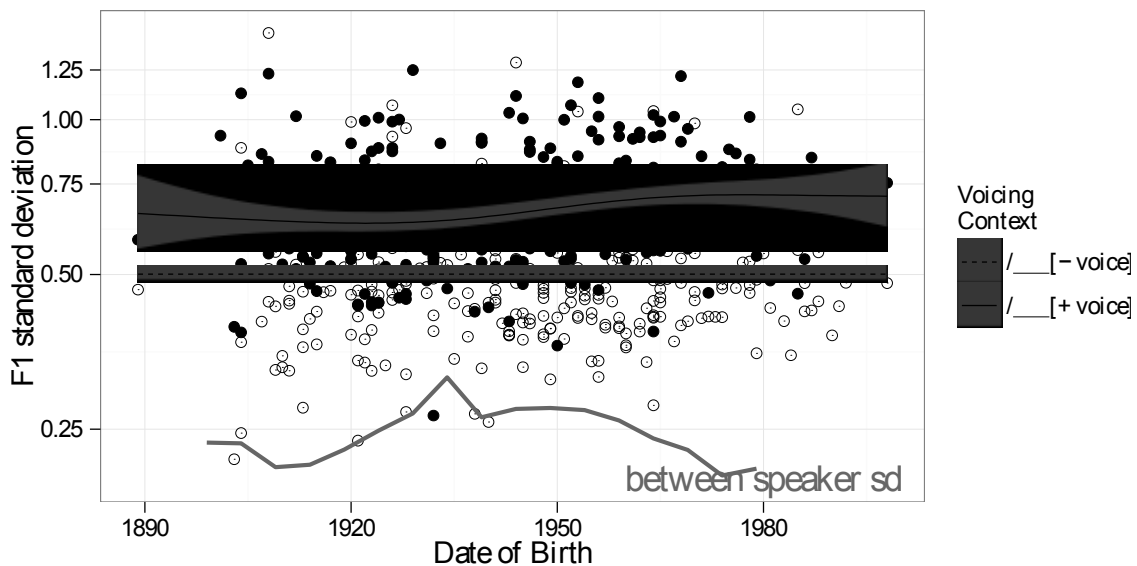


Figure 6 Comparison of within-speaker variation (one point per speaker) to between-speaker variation (grey line). The smoothing lines are penalized cubic regression splines. Between-speaker standard deviation calculated based on a moving 20 year date of birth window at 5 year increments. The y-scale is logarithmic.

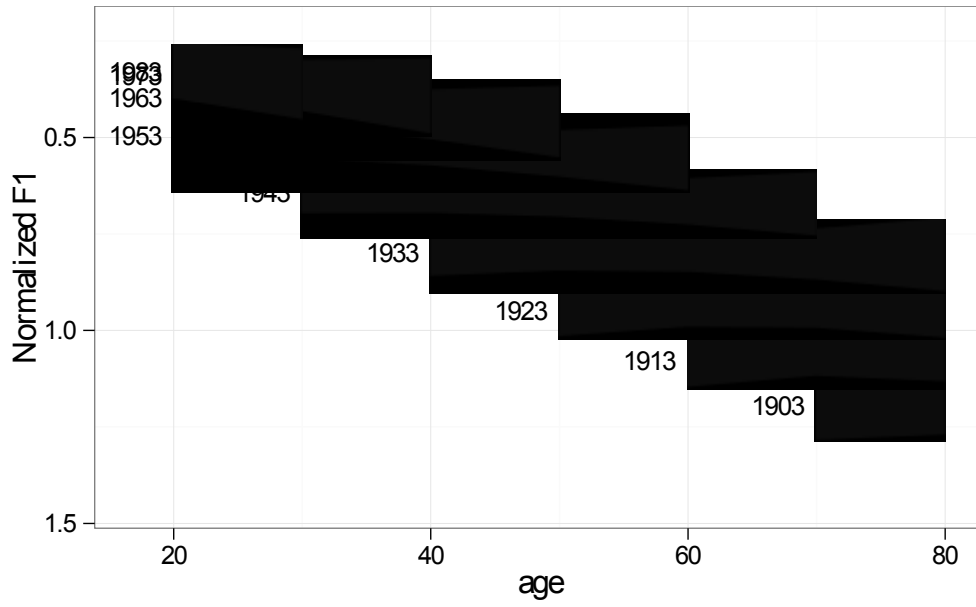


Figure 7 Predicted F1 of pre-voiceless /ay/ for different Date of Birth cohorts at different ages. `ages.gam(NormalizedF1 ~ ti(age) + ti(dob) + ti(age,dob))`



Figure 8 /ay/ glide height across the 20th century. The smoothing lines are penalized cubic regression splines. Dark grey lines correspond to /d/; light grey lines correspond to /t/. Solid lines correspond to faithful (non-flapping) realizations of /t, d/; dashed lines correspond to flapped realizations of /t, d/.

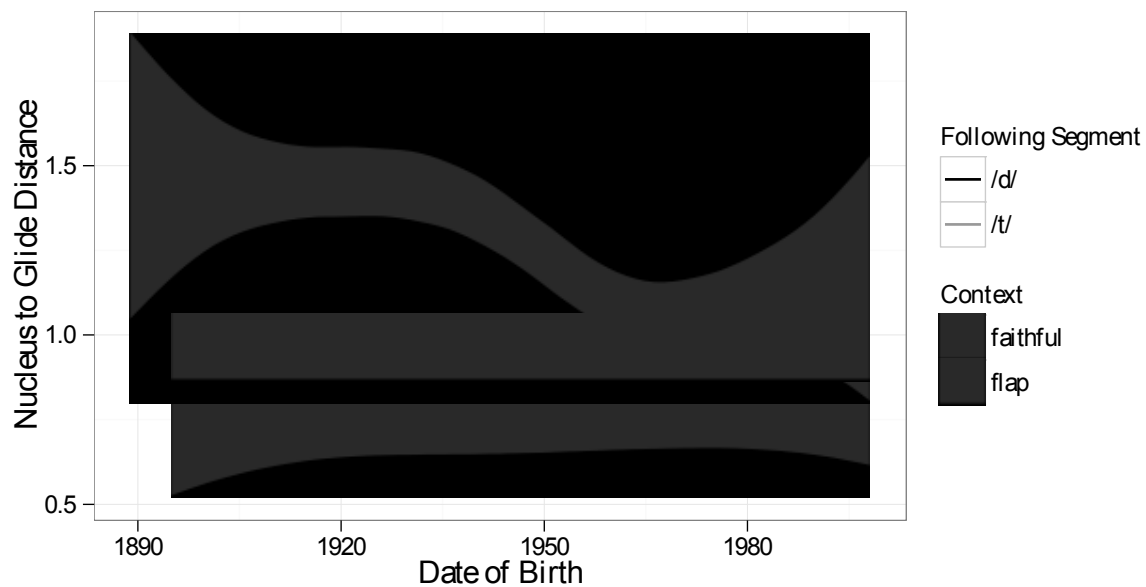


Figure 9 Distance from /ay/ nucleus (maximum F1) to glide (80% of vowel duration) across the 20th century. The smoothing lines are penalized cubic regression splines. Dark grey lines correspond to /d/; light grey lines correspond to /t/. Solid lines correspond to faithful (non-flapping) realizations of /t, d/; dashed lines correspond to flapped realizations of /t, d/.

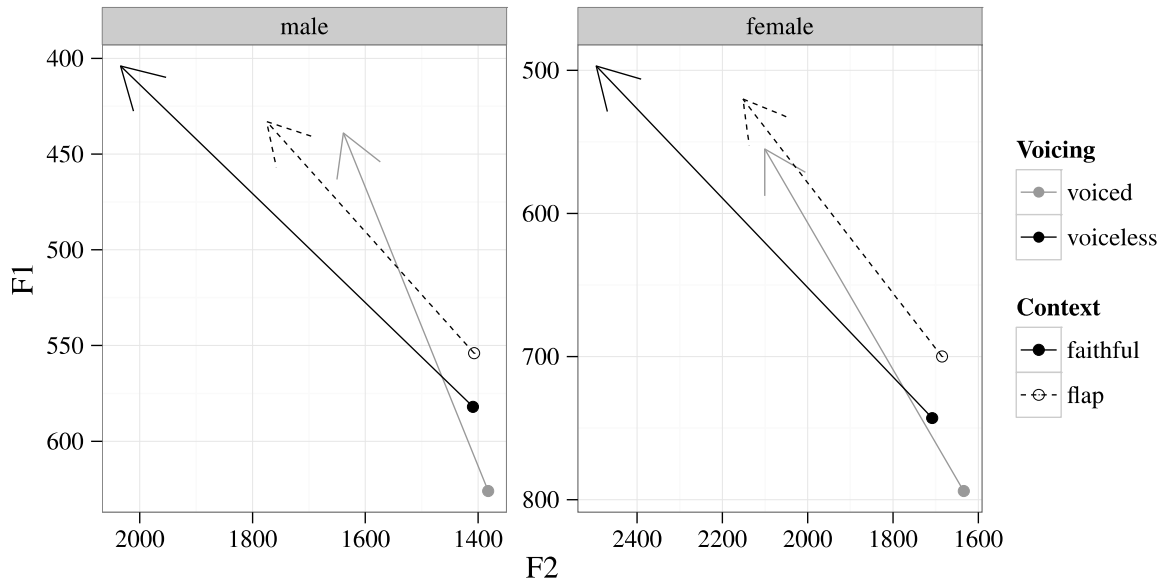


Figure 10 Nucleus and glide measurements of pre-voiced, pre-voiceless, and pre-/t/-flap /ay/ from Victoria, B.C. Data from Rosenfelder (2005).

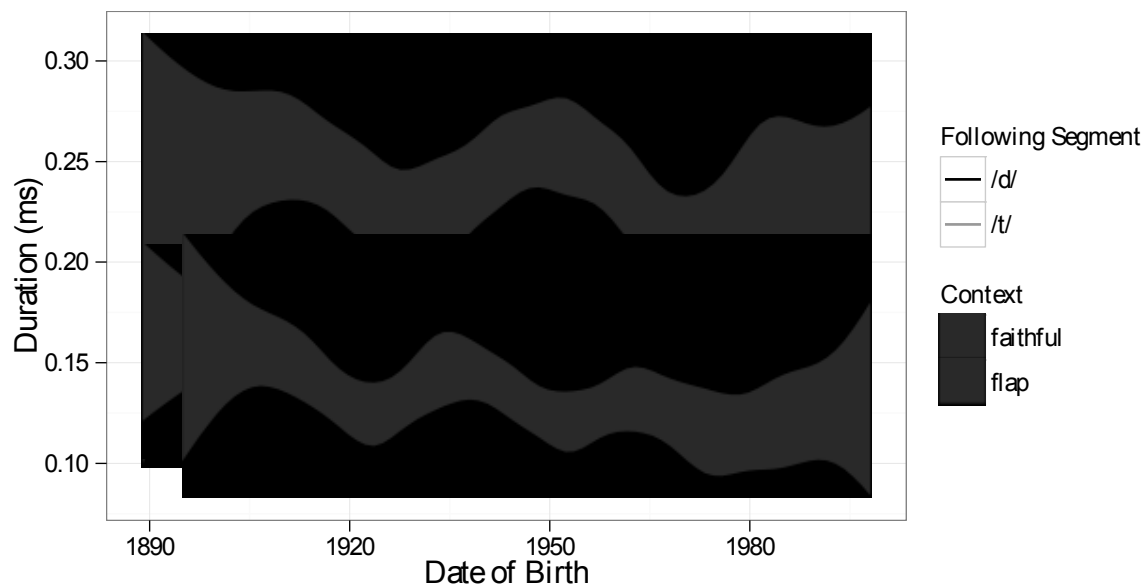


Figure 11 Diachronic trajectory for /ay/ duration across the 20th century. The smoothing lines are penalized cubic regression splines. Dark grey lines correspond to /d/; light grey lines correspond to /t/. Solid lines correspond to faithful (non-flapping) realizations of /t, d/; dashed lines correspond to flapped realizations of /t, d/.

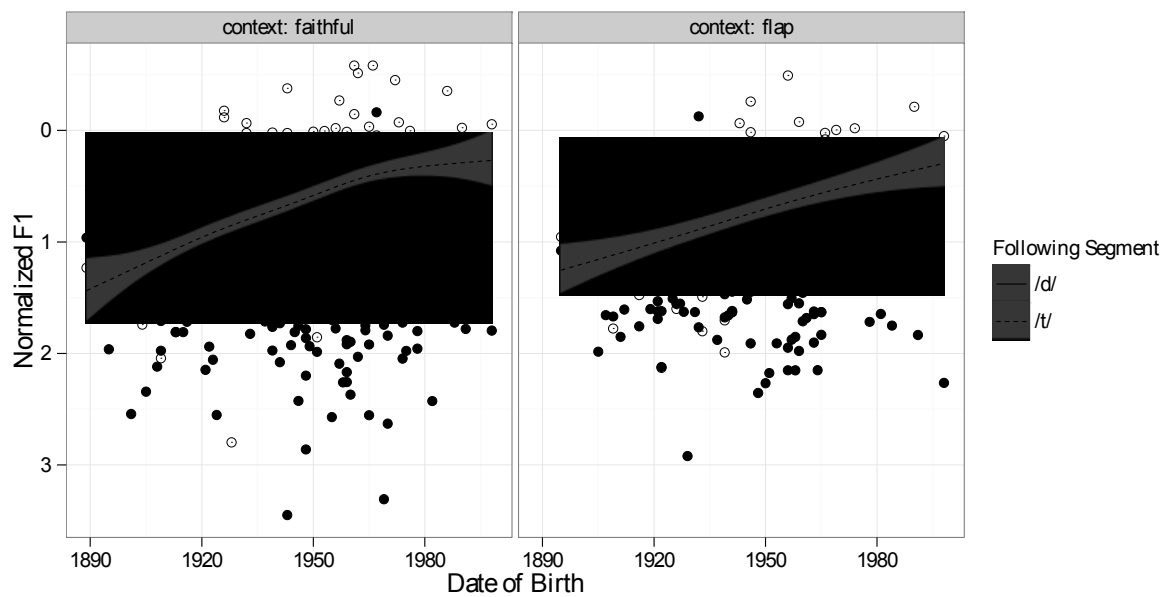


Figure 12 Speaker F1 means for /ay/ preceding different /t/ and /d/ realizations.

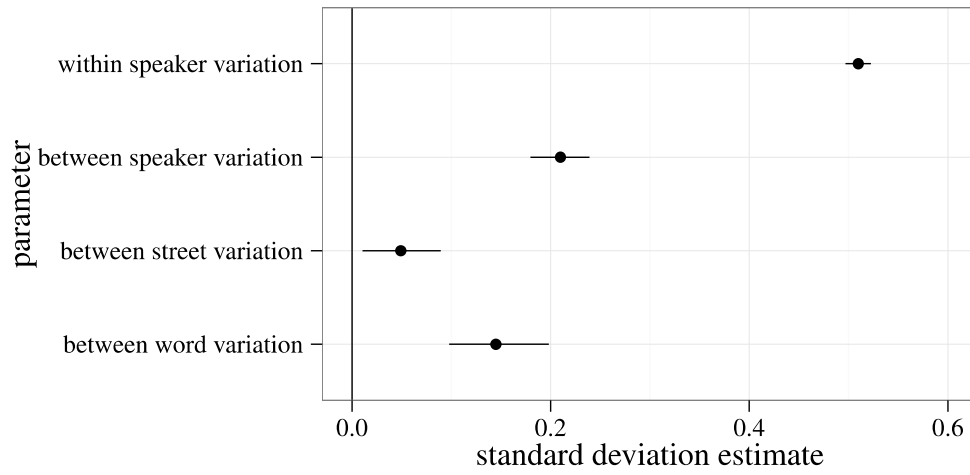


Figure 13 Estimated standard deviations from the Stan model. Intervals represent the range of the 95% Highest Posterior Density Interval.

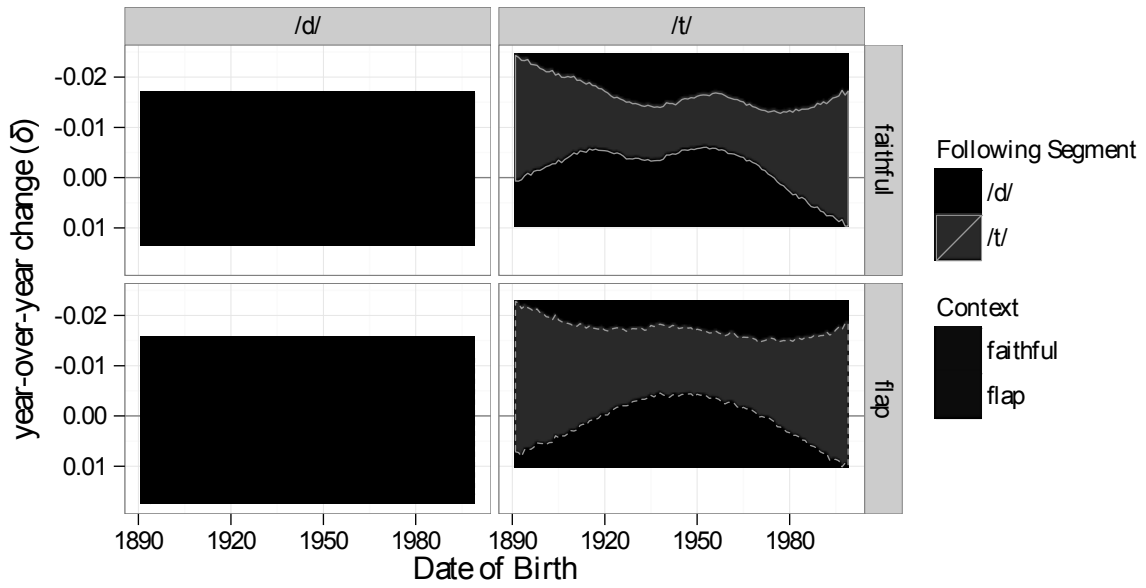


Figure 14 Model estimates and 95% credible intervals for the year-over-year differences (δ) i.e. the rate of change. Dark grey lines correspond to */d/*; light grey lines correspond to */t/*. Solid lines correspond to faithful (non-flapping) realizations of */t, d/*; dashed lines correspond to flapped realizations of */t, d/*.

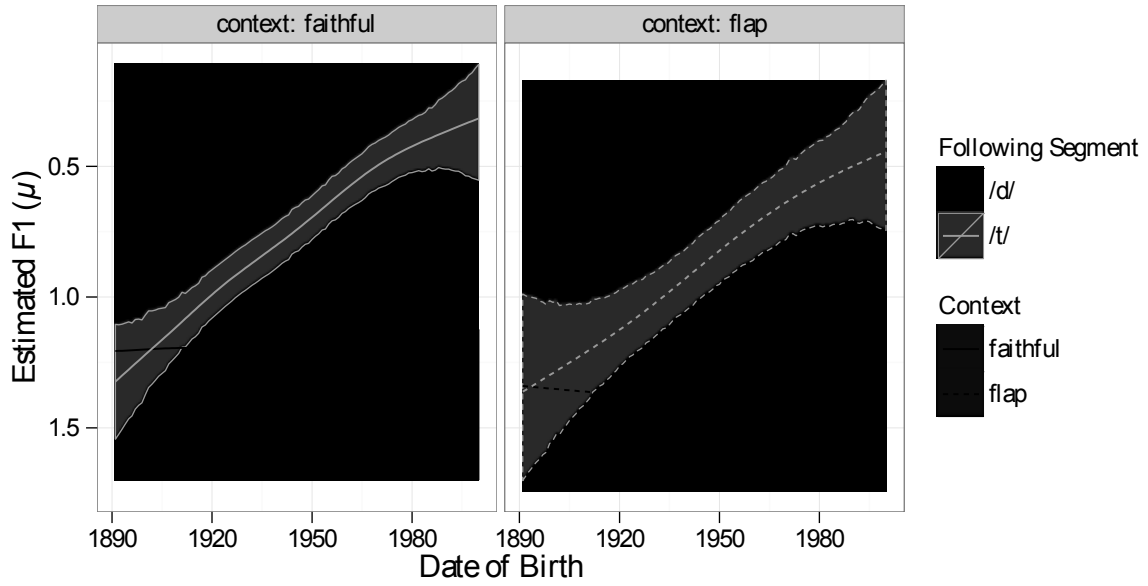


Figure 15 Model estimates and 95% credible intervals for normalized /ay/ F1. Dark grey lines correspond to /d/; light grey lines correspond to /t/. Solid lines correspond to faithful (non-flapping) realizations of /t, d/; dashed lines correspond to flapped realizations of /t, d/.

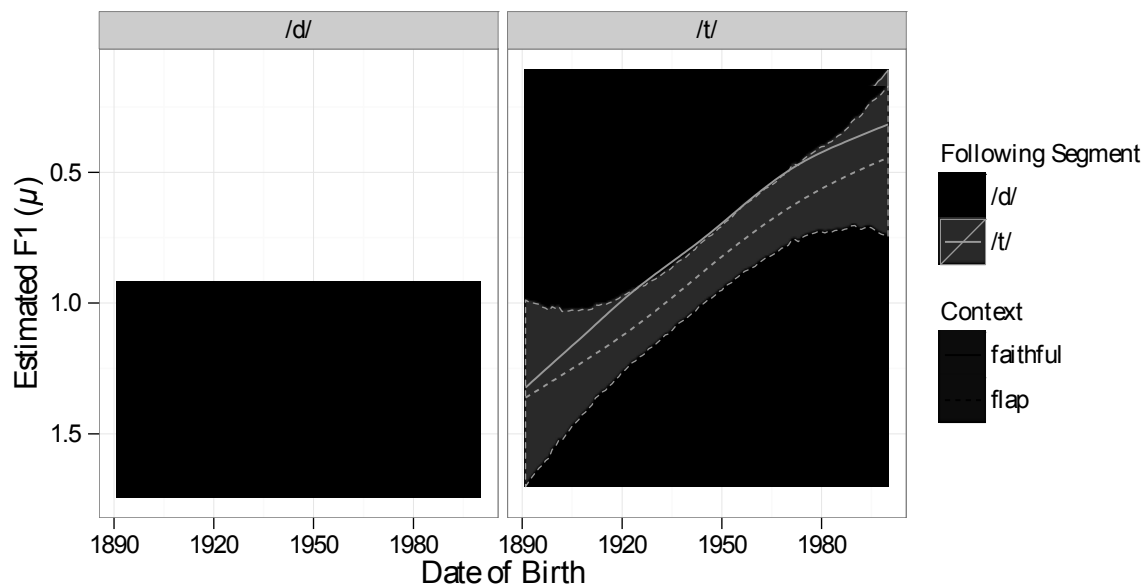


Figure 16 Model estimates and 95% credible intervals for normalized /ay/ F1. Dark grey lines correspond to /d/; light grey lines correspond to /t/. Solid lines correspond to faithful (non-flapping) realizations of /t, d/; dashed lines correspond to flapped realizations of /t, d/.

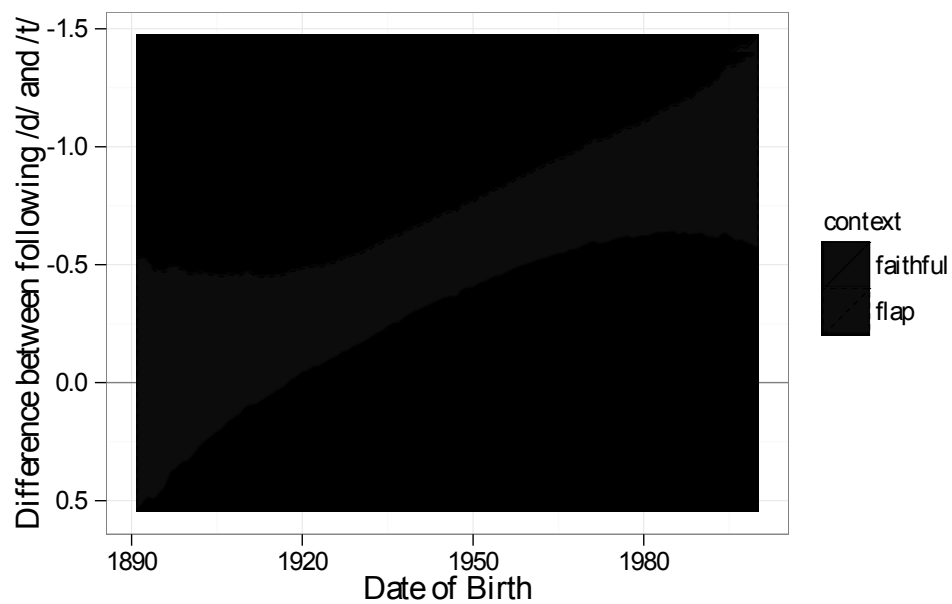


Figure 17 Estimated /ay/ height difference between pre-/t/ and pre-/d/ contexts. Solid lines correspond to faithful (non-flapping) realizations of /t, d/; dashed lines correspond to flapped realizations of /t, d/.

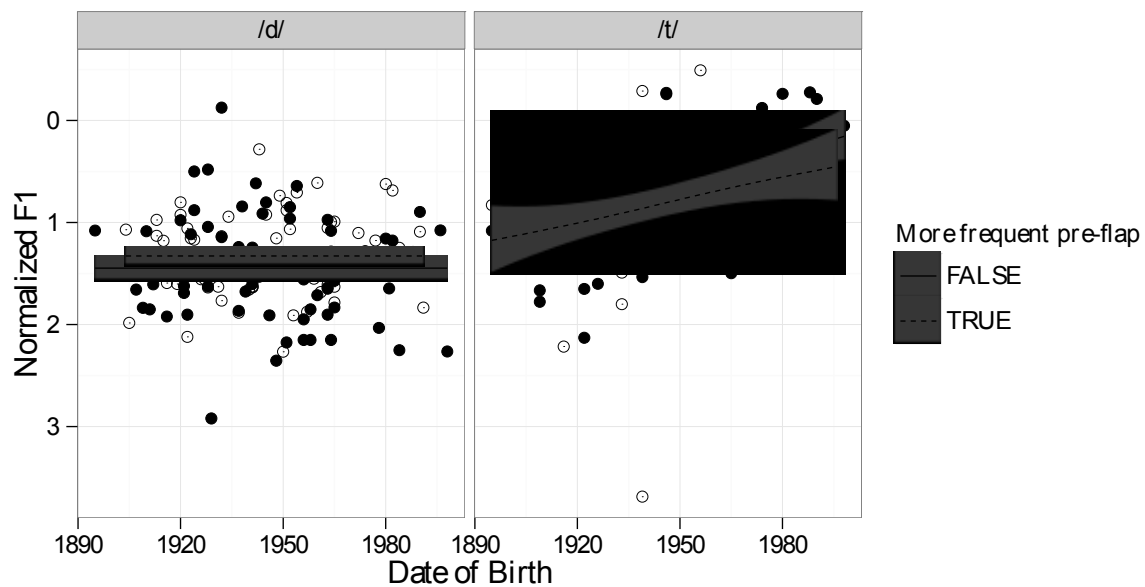


Figure 18 Estimated F1 of /ay/ in flapping contexts, divided by roots which appear more often in flapping contexts, and those which appear more often in faithful contexts. The smoothing lines are penalized cubic regression splines.

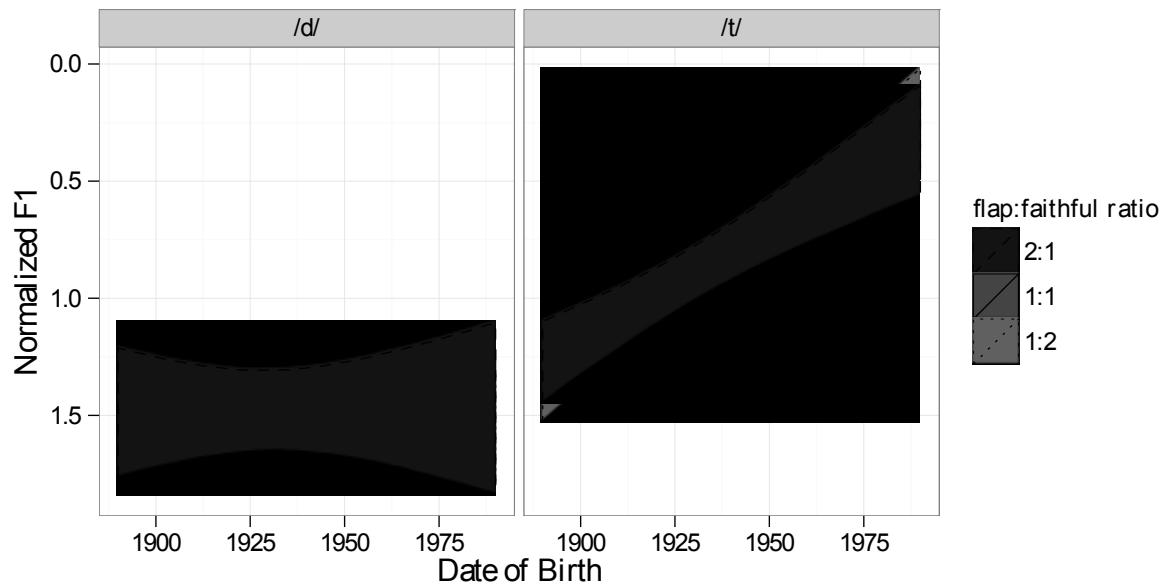
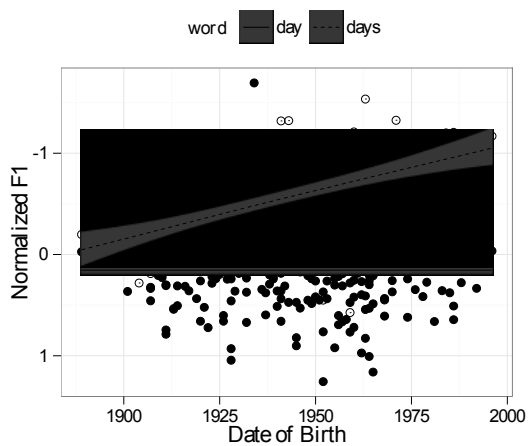


Figure 19 95% percentiles of the fitted values from 10,000 bootstrap replicates.

a)



b)

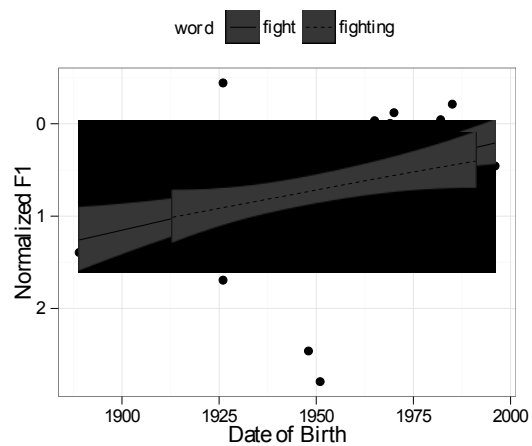
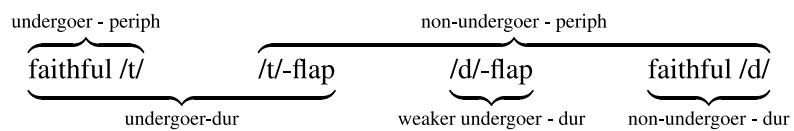


Figure 20 Mean F1 over date of birth for different lexical items. The smoothing lines are penalized cubic regression splines.



Example (5)

TABLES

Table 1 Parameter estimates and derived values from the mixed effects model $\text{NormalizedF1} \sim \text{Decade} * \text{following voicing} * \log_2\text{Freq} + (\text{Decade} \text{Word}) + (\text{following voicing} \text{Speaker})$. Decade is the speakers' date of birth, centered at 1950, and divided by 10. $\log_2\text{freq}$ is $\log_2(\text{frequency}) - \text{median}(\log_2(\text{frequency}))$, based on frequencies from SUBTLEX _{US} . The 95% confidence intervals are based on 5,000 semiparametric bootstrap replicates fitted by <code>bootMer</code> from <code>lme4</code> v1.1-7. The density distributions of the bootstraps are provided in the final column.	71
Table 2 Comparisons of models that differ in terms of how word frequency was included.	72
Table 3 Random effects' standard deviations and correlations from the model described in Table 1.....	73
Table 4 Number of tokens preceding flapped and faithful /t/ and /d/.	74
Table 5 Revised number of tokens preceding flapped and faithful /t/ and /d/.	75
Table 6 Mean duration of /ay/ preceding both faithful and flapped /t/ and /d/.	76
Table 7 Duration differences between /t/ and /d/ in faithful and flapping contexts.	77
Table 8 Estimated participation rates of /ay/ raising on the basis of offglide-peripheralization. .	78
Table 9 Estimated participation rates of /ay/ raising on the basis of phonetic duration.	79
Table 10 SUBTLEX _{US} frequency norms for the root unite.	80
Table 11 Summed frequency norms for the root unite, and the ratio of flapping frequency to faithful frequency.	81
Table 12 Parameter estimates and derived values from the mixed effects model $\text{NormalizedF1} \sim \text{Decade} * \text{following voicing} * \log_2\text{Ratio} + (\text{Decade} \text{Word}) + (\text{following voicing} \text{Speaker})$. Decade is the speakers' date of birth, centered at 1950, and divided by 10. $\log_2\text{ratio}$ is $\log_2(\text{flap}/\text{faithful})$, based on the frequencies described in §4.4. The 95% confidence intervals are based on 10,000 semiparametric bootstrap replicates fitted by <code>bootMer</code> from <code>lme4</code> v1.1-7. The density distributions of the bootstrapped parameters are provided in the final column.	82
Table 13 Comparison of models that differ in terms of how the ratio of flaps:faithful was included.	83

Parameter	Estimate	95% CI	Bootstrap
Intercept*	1.442	[1.404, 1.48]	
Decade	-0.005	[-0.017, 0.007]	
Doubling Frequency*	-0.013	[-0.022, -0.003]	
Decade × Frequency	-4×10^{-4}	[-0.002, 0.002]	
[-voice]*	-0.753	[-0.829, -0.679]	
Decade × [-voice]*	-0.106	[-0.129, -0.083]	
Doubling Frequency × [-voice]	3×10^{-4}	[-0.018, 0.019]	
Decade × Doubling Frequency × [-voice]	1×10^{-4}	[-0.004, 0.004]	

Table 1 Parameter estimates and derived values from the mixed effects model `NormalizedF1 ~ Decade * following voicing * log2Freq + (Decade|Word) + (following voicing|Speaker)`. Decade is the speakers' date of birth, centered at 1950, and divided by 10. `log2freq` is $\log_2(\text{frequency}) - \text{median}(\log_2(\text{frequency}))$, based on frequencies from SUBTLEX_{US}. The 95% confidence intervals are based on 5,000 semiparametric bootstrap replicates fitted by `bootMer` from `lme4` v1.1-7. The density distributions of the bootstraps are provided in the final column.

	df	AIC	BIC	logLik	deviance	Chisq	Chi Df	p
no freq	11	164234.1	164336.8	-82106.03	164212.1			
+freq	12	164226.2	164338.3	-82101.10	164202.2	9.848	1	0.002
voicing×freq	13	164228.2	164349.7	-82101.10	164202.2	0.000	1	0.999
decade×voicing×freq	15	164228.2	164372.2	-82101.02	164202.0	0.167	2	0.920

Table 2 Comparisons of models that differ in terms of how word frequency was included.

Group	Name	Std.Dev	Corr
Word	(Intercept)	0.24	
	Decade	0.03	0.49
Speaker	(Intercept)	0.12	
	[-voice]	0.2	-0.58
Residual		0.63	

Table 3 Random effects' standard deviations and correlations from the model described in Table 1.

Following Segment	faithful	flap
/d/	626	208
/t/	2,392	336

Table 4 Number of tokens preceding flapped and faithful /t/ and /d/.

Following Segment	faithful	flap
/d/	626	208
/t/	2,392	285

Table 5 Revised number of tokens preceding flapped and faithful /t/ and /d/.

Following Segment	Context	Mean Duration (ms)	Difference from next shortest
/t/	flap	135	
/t/	faithful	146	11
/d/	flap	179	32
/d/	faithful	228	49

Table 6 Mean duration of /ay/ preceding both faithful and flapped /t/ and /d/.

context	/ay/ pre-/d/	ay pre-/t/	difference
faithful	219	135	84
flap	161	124	36

Table 7 Duration differences between /t/ and /d/ in faithful and flapping contexts.

Following Segment	Context	Nucleus to glide distance	Participation
/t/	faithful	1.44	1.00
/t/	flap	0.91	0.10
/d/	flap	0.87	0.04
/d/	faithful	0.85	0.00

Table 8 Estimated participation rates of /ay/ raising on the basis of offglide-peripheralization.

Following Segment	Context	Mean Duration (ms)	Participation
/t/	faithful	135	1.13
/t/	flap	146	1.00
/d/	flap	179	0.60
/d/	faithful	228	0.00

Table 9 Estimated participation rates of /ay/ raising on the basis of phonetic duration.

root	word	freq	context
unite	unite	3.02	faithful
unite	reunite	0.71	faithful
unite	unites	0.53	faithful
unite	reunites	0.04	faithful
unite	united	50.27	flap
unite	reunited	1.78	flap
unite	uniting	0.29	flap
unite	reuniting	0.27	flap

Table 10 SUBTLEX_{US} frequency norms for the root unite.

root	faithful	flap	flap:faithful
unite	4.3	52.61	12.23

Table 11 Summed frequency norms for the root unite, and the ratio of flapping frequency to faithful frequency.









Parameter	Estimate	95% CI	Bootstrap
Intercept*	1.487	[1.322, 1.65]	
Decade	0.005	[-0.038, 0.047]	
log ₂ flap ratio	-0.009	[-0.069, -0.049]	
Decade × Ratio	-0.007	[-0.027, 0.013]	
[-voice]*	-0.793	[-0.999, -0.592]	
Decade × [-voice]*	-0.107	[-0.16, -0.054]	
Ratio × [-voice]	0.023	[-0.047, 0.096]	
Decade × Ratio × [-voice]	0.015	[-0.009, 0.038]	

Table 12 Parameter estimates and derived values from the mixed effects model `NormalizedF1 ~ Decade * following voicing * log2Ratio + (Decade|Word) + (following voicing|Speaker)`. Decade is the speakers' date of birth, centered at 1950, and divided by 10. `log2ratio` is $\log_2(\text{flap}/\text{faithful})$, based on the frequencies described in §4.4. The 95% confidence intervals are based on 10,000 semiparametric bootstrap replicates fitted by `bootMer` from `lme4` v1.1-7. The density distributions of the bootstrapped parameters are provided in the final column.

	df	AIC	BIC	logLik	deviance	Chisq	Chi Df	p
no ratio	11	660.161	705.628	-319.081	638.161			
+ratio	12	662.130	711.731	-319.065	638.130	0.031	1	0.859
voicing×ratio	13	664.124	717.731	-319.062	638.124	0.006	1	0.938
decade×voicing×ratio	15	666.260	728.261	-318.130	636.260	1.864	2	0.394

Table 13 Comparison of models that differ in terms of how the ratio of flaps:faithful was included