An Acoustic Study of Canadian Raising in Three Dialects of North American English

by

D. Sky Onosson
B.A., University of Manitoba, 1994
M.A., University of Manitoba, 2010

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Linguistics

Supervisory Committee

An Acoustic Study of Canadian Raising in Three Dialects of North American English

by

D. Sky Onosson
B.A., University of Manitoba, 1994
M.A., University of Manitoba, 2010

Dr. Sonya Bird (Department of Linguistics)
Co-Supervisor

Dr. Alexandra D'Arcy (Department of Linguistics)
Co-Supervisor

Dr. Josef Fruehwald (School of Philosophy, Psychology and Language Sciences, The University of Edinburgh)
Outside Member

Abstract

Supervisory Committee

Dr. Sonya Bird (Department of Linguistics)
Co-Supervisor

Dr. Alexandra D'Arcy (Department of Linguistics)
Co-Supervisor

Dr. Josef Fruehwald (School of Philosophy, Psychology and Language Sciences, The University of Edinburgh)
Outside Member

"Canadian Raising" (CR) is a phonological process typical of Canadian English, defined as the production of /aj, aw/ with raised nuclei before voiceless codas, e.g. in *about*. This dissertation investigates the relationship between CR and another process which abbreviates vowels in the same phonological context in most English dialects: pre-voiceless vowel abbreviation (PVVA). This study sampled three North American dialects: Canada, and the American West and North. Comparisons of vowel duration and formant trajectories revealed common patterns and specific differences between these dialects related to both CR and PVVA. Comparisons of vowel formant trajectories were conducted using statistical techniques for comparing curvilinear datasets, employed in novel methodology which utilizes multiple models of time-scaling. Results indicate that the allophonic production of /aw/ differs in Canadian English in relation to the other dialects, while /aj/ follows a common pattern in all three. I argue that PVVA is achieved through the gestural reorganization of vowels preceding voiceless coda, with the dynamic nature of diphthongs making possible several patterns of abbreviation, two of which are attested in these data: truncation of the onset i.e. the diphthongal nucleus, and compression of the overall trajectory; truncation of the offset is also attested for some monophthongs. Differences in selection of which of these abbrevatory patterns applies to /aw/ in Canadian English versus other dialects accounts for the observed differences in phonetic output. These results indicate that it is worth reconsidering several aspects of the current conception of CR, as follows. First, diphthong-raising processes can be directly linked to

the more common process of vowel abbreviation, with consideration of how diphthongal gestures are organized, and reorganized in relation to post-vocalic voicing gestures. Second, that /aw/-raising appears to be distinctly Canadian. And third, that /aj/-raising is not specifically Canadian, suggesting that the two terms be described and named distinctly. This dissertation contributes to the literature on sociophonetics in two major ways: by indicating how CR is directly connected to PVVA in contemporary speech, beyond their surmised historical connections; and, by developing novel methodology for the analysis of dynamic formant trajectories, involving comparison of different time-scaling methods.

Table of Contents

List of Tables

List of Figures

Acknowledgments

This dissertation was produced while the author was enrolled at the University of Victoria, British Columbia. Data collection for the primary Canadian study was carried out in Winnipeg, Manitoba. Both communities inhabit the traditional territory of several indigenous peoples. The University of Victoria has issued the following statement with respect to its relationship with those groups: “We acknowledge and respect the Lkwungen-speaking peoples on whose traditional territory the university stands and the Songhees, Esquimalt and WSÁNEĆ peoples whose historical relationships with the land continue to this day,” (University of Victoria, 2017). Winnipeg is located on Treaty 1 territory, traditional territory of Anishinaabeg, Cree, Oji-Cree, Dakota, and Dene peoples, and the homeland of the Métis Nation (Canadian Association of University Teachers 2016).

Data collection in the United States was conducted in Denver, Colorado, and Madison, Wisconsin. Indigenous cultures inhabiting the region comprising present-day Colorado have included the Ancestral Puebloans (Anasazi), Frémont, Ute, Apache, Navajo, Cheyenne, Comanche and Arapaho (see §2.4.1). In Wisconsin, their counterparts have included the Ho-Chunk (Winnebago), Huron, Chippewa (Ojibwe), Sauk (Sac), Fox, Miami, and Menominee (see §2.5.1). Neither of these lists should be taken as exhaustive nor definitive.

Work for this dissertation was conducted within the following free (as in beer) software environments:

Praat (Boersma & Weenink, 2016)

R Programming Language (R Core Team, 2016)

RStudio IDE (RStudio Team, 2016)

In the words of the inimitable John 'Josco' Scoles:

"You're all good people."

Dedication

I would like to dedicate this dissertation to my two children, Shoki Kriyah Onosson and Cyan James Onosson, for putting up with periodic mental absences of their father during the work involved in creating it. I would also like to sincerely dedicate it to my wife Annika Shawoki James Onosson, who suffered a severe medical condition and became hospitalized towards the end of the final stage of my degree. I could not have even attempted to carry out the work of either my M.A. or Ph.D. without her full support and hard work, even before having to deal with this situation beyond her control; and even in the face of that she helped ensure that I could find time to complete writing this so that I could find a way to properly take care of both her and our children into the future. The gratitude I owe her for everything she has done, and the respect I feel towards her for dealing with everything that life has put in front of her, cannot be adequately expressed in any words I could write here. The greatest sacrifice made in preparing this document has truly been hers.

Chapter 1 Introduction

"… the category 'diphthong' cannot be defined by the presence or absence of some necessary and sufficient conditions of membership. Instead, it is necessary to find a series of features that contribute in different degrees …"

— Fernando Sanchez Miret (1998, p. 37)

Canadian Raising is a well-known stereotypical feature of Canadian English, referring to the articulation of the diphthongs /aj, aw/ with raised nuclei when occurring before voiceless codas within the same phonological foot (Chambers 1973; Paradis 1980; inter alia). In the earliest known account of Canadian Raising as a singular phonological process, Martin Joos (1942) opined that "[t]he starting-point for this articulatory difference was presumably the relative shortness of English vowels before fortis [i.e. voiceless] consonants," (p. 142). Here, Joos makes reference to a pattern of vowel duration whose conditioning environment matches that of Canadian Raising: longer vowels occur before tautosyllabic voiced codas, and shorter vowels before voiceless codas. This pattern of pre-voiceless shortening or abbreviation is commonplace in many or perhaps all dialects of English, and is quite well-documented beyond Canada (Heffner 1937; Peterson & Lehiste 1960; inter alia). In light of the presumed historical relationship between Canadian Raising and pre-voiceless vowel abbreviation (Joos 1942; Chambers 1973; Gregg 1973; inter alia) and their identical conditioning environment, an important question to be asked is: *What is* *the contemporary role of duration in the production of Canadian Raising?* Joos was of the opinion that while pre-voiceless abbreviation was the historical source for Canadian Raising, it had since become a "secondary" aspect of the diphthongs for Canadian English speakers. A few studies have discussed the role of vowel abbreviation in the context of raising of the diphthong /aj/ (Myers 1997; Moreton & Thomas 2007). However, to my knowledge there exists to date only one acoustic study of Canadian Raising in a Canadian speech context which includes and compares both of the relevant diphthongs, and which also incorporates durational differences. This is documented in Hall (2016a,b), who investigated Canadian Raising among Toronto and Vancouver speakers. Hall's primary analytic technique involves the *smoothing spline analysis of variance* (SSANOVA; described in more detail in §4.4.2). Under typical applications, "[t]he SS-ANOVA method normalizes vowel duration across tokens and therefore excludes this timing information," (Hall 2016b:6). Such was the case in Hall's use of the technique as well, although she did separately report distinct durational patterns between Canadian Raising diphthongs in her sample populations; these results are included in the discussion of vowel duration (production) patterns in §2.1.1. This dissertation seeks to contribute to our knowledge in the area of research on Canadian Raising by adapting the SSANOVA technique, and another statistical method for working with non-linear data, *generalized additive mixed models*, or GAMMs, to incorporate durational differences between the compared groups; in this case, vowels/diphthongs in abbrevatory vs. non-abbrevatory contexts.

In the first acoustic study of vowel production conducted in Winnipeg, Hagiwara (2006) posed the following question for future researchers to take up: "How do raising and non-raising dialects differ with respect to the effects of the voicing/lengthening correlation?” (p. 138). This dissertation seeks to address Hagiwara’s question through an investigation of how duration is implicated in the production of Canadian Raising. As noted above, vowel abbreviation in pre-voiceless context occurs widely throughout the English-speaking world, whereas Canadian Raising is much more restricted. Even if the two processes are connected contemporaneously in Canadian speech, the specific role that durational abbreviation plays within Canadian Raising is particular to Canadian English, and may differ from that of other dialects which lack Canadian Raising yet possess pre-voiceless vowel abbreviation. In order to answer Hagiwara’s question, this dissertation looks beyond the set of Winnipeg subjects, adding two distinct sets of speakers of American dialects which do not exhibit canonical Canadian Raising, i.e. pre-voiceless raising of both /aj/ and /aw/: Denver, Colorado representing *The West*; and Madison, Wisconsin representing *The North* (Labov et al. 2006). The two American dialects chosen for comparison were selected for their broad similarity to Canadian English in terms of vowel production, but with speakers in Denver expected not to exhibit raising of either Canadian Raising diphthong (per Labov et al. 2006), and speakers in Madison known to potentially exhibit raising of /aj/, but not expected to have raising of /aw/ (Labov et al. 2006, Purnell 2010).

Above, I noted that canonical Canadian Raising involves the raising of /aj/ and /aw/ before voiceless codas. Raising of /aj/ alone has been described among speakers in several regions of the United States both contiguous with the Canadian border (Vance 1987; Allen 1989; Dailey-O-Cain 1997; Niedzielski 1999; Roberts 2007; inter alia) and non-contiguous (Greet 1931; Allen 1989; Moreton & Thomas 2007; Fruehwald 2013; Carmichael 2015; Davis, Berkson & Strickler 2016; inter alia), and similar processes have been described in varieties of English spoken outside North America as well (Gregg 1973; Trudgill 1986; Britain 1997; inter alia). In such non-Canadian contexts, the term *Canadian Raising* is typically used to refer to /aj/-raising, despite the absence of concomitant /aw/-raising. Considering the population differences between the United States and Canada, it would not be surprising if the largest population of North American /aj/-raisers turned out to consist of mainly U.S. speakers, meaning that the most distinctive aspect of Canadian Raising within Canada concerns its occurrence in /aw/. And because use of the term Canadian Raising in the U.S. almost always solely refers to /aj/-raising, we have an interesting and potentially confusing situation where the term *Canadian* Raising is often used to refer solely to its least distinctively Canadian aspect, namely raising of /aj/.

At the same time, doubts have been raised by prominent researchers of Canadian English such as Chambers (1973, 1989) and Boberg (2008) on the appropriateness of describing the allophonic raising patterns of both /aj, aw/ as a unitary phenomenon, as the two diphthongs exhibit distinct characteristics even in Canadian English, such as variable fronting of the nucleus of /aw/. The occurrence of /aj/-raising apart from /aw/-raising, the otherwise distinctiveness of Canadian /aj/ and /aw/, and the disjointed usage of the term Canadian Raising within the literature all point to another important question: *What is the most apt characterization of Canadian Raising?* Is it raising of /aj/, raising of /aw/, or raising of both? Are there other phonetic qualities aside from nuclear height which are significant, such as vowel duration, and should be included as well?

To summarize, this dissertation thus addresses the two research questions posed above, *What is the contemporary role of duration in the production of Canadian Raising?* and *What is the most apt characterization of Canadian Raising?* through an examination of the acoustic differences between three North American English dialects with varying patterns of diphthong-raising, and the incorporation of durational abbreviation patterns into an analysis of formant frequency trajectories of the diphthongs /aj, aw/. The layout of this dissertation is as follows. Chapter 2 reviews the existing literature describing research on the topic of vowel duration in English, Canadian Raising itself, and the phonological and phonetic characteristics of vowels in each of the three cities where the studies were carried out. Chapter 3 describes the methodology used in carrying out each acoustic study, from design to recording to analysis. Chapter 4 discusses the results from analysis of each dataset, with separate sections on acoustic vowel positions, vowel duration patterns, diphthong positions and trajectories, and statistical methods of comparing diphthong trajectories from allophones with significantly different durations. Chapter 5 synthesizes the information presented in Chapter 4 into a response to the research questions central to the dissertation, providing a description of Canadian Raising which incorporates the role of allophonic durational abbreviation while recognizing the distinct patterning of each of the diphthongs /aj, aw/, both in Canadian English as well as in related but divergent American English dialects.

1.1 A NOTE ON THE TRANSCRIPTION OF DIPHTHONGS

Several different methods for transcribing diphthongs are utilized in the phonetic and phonological literature. The main point of difference concerns the notation of the off-glide portion of the articulation, which may be indicated by a glide e.g. <j, w> or by a vowel. For the latter, there is also a distinction made which especially concerns English diphthongs (here focusing strictly on North American varieties) with regard to the quality of that vowel, with some selecting a lax vowel <ɪ, ʊ>, others a tense one <i, u>. In the phonetics literature, the use of lax vowels is fairly common; for example, Ladefoged (2006) has [aɪ, aʊ, ɔɪ]. In the phonological literature, the forms /aj, aw, ɔj/ are often preferred; for example, these are the forms used by Hammond (1999) with the exception that Americanist /y/ is used in place of IPA /j/. Although not stated explicitly, this may be because the use of glide symbols makes explicit their phonological distinction from the vocalic nucleus, whereas multiple vowel symbols are phonologically ambiguous with respect to the location of the nucleus. Sociophonetic notations are notably varied, and include all three possibilities of lax vowel, tense vowel, or glide symbol for the off-glide.

In this dissertation, I generally use /aj, aw, ɔj/ except when referencing other sources, where the original form is adhered too. The phonetic qualities of the diphthongs will be presented in a variety of ways apart from transcription, focusing on their acoustic qualities—most importantly formant values, but also duration—and discussing these visually and/or statistically rather than merely notationally. The use of phonological notation allows such phonetic details to be deliberately obscured when discussing the diphthongs from a more abstract, and therefore more general viewpoint; for example, when discussing phonemes in the context of allophonic or dialectal variation. In Chapter 6, the topic of diphthong notation will be revisited, and some proposals made for the most appropriate phonological and phonetic notations for each of the dialects investigated in this study.

1.2 A NOTE ON ETHNOLINGUISTIC DIFFERENCES

This dissertation presents data on speakers representing samples of local populations in three communities: Winnipeg, Manitoba; Denver, Colorado; and Madison, Wisconsin. During recruitment and subsequent field interviews, no attempt was made to restrict or segregate speakers based on actual or perceived ethnic background or other social group affiliation, aside from geographic locale, both during childhood and at the time of the interview. Based solely on my own recall of participants' physical appearances, which cannot be taken as definitive in any respect, the majority of participants would probably fall into the poorly-defined category of "white" (bearing in mind that their own self-identification may or may not agree with this assessment), with no more than one or two exceptions in each location.

Linguistic differences between ethnic groups have been well-documented for a wide variety of regions and languages. Examples of this abound throughout the sociolinguistic literature; one especially topical example with respect to this dissertation is Boberg (2005; also discussed in Boberg 2010) which documents ethnolinguistic differences between several long-established groups residing in Montreal, Canada. Although Labov et al. (2006), an important source of background information on the dialect regions involved in this dissertation, gathered demographic speaker data during recruitment, their participants were not restricted to any particular ethnic group or groups, nor excluded on any such basis. By far the largest ethnic group within their overall sample are of reported German ancestry, at 28.5%; the second-largest group is undifferentiated "white" at 10.5%. However, only one group is singled out by Labov et al. for its own chapter and discussion, African-

Americans. As such, the overall conclusions reached by Labov et al. in the rest of that study may be taken as largely pertaining to "white" North Americans, understood broadly.

Focusing on Canada, within Labov et al.'s Canadian sample (n=38), the largest ethnic group was Scots-Irish, at 29%, with only a single Canadian individual self-identifying as "white" (one suspects this may speak to differences in how racial categories are perceived in the United States vs. in Canada). In a discussion of ethnolinguistic differences in Canada, Boberg (2010) cites and discusses only the 2005 Montreal study mentioned above. With respect to Winnipeg in particular, the only extant published study, Hagiwara (2006), explicitly makes no attempt "to control for possible Winnipeg-internal ethnic, geographic, or cultural dialectal variants," (p. 128). None of this is to say that ethnolinguistic differences do not exist in Canadian regions outside of Montreal, of course. Hoffman & Walker (2010) examined two variable sociophonetic features in ethnic communities in Toronto, finding that speakers differed in rate of usage of ethnically-associated forms depending on individual factors related to group affiliation. In Winnipeg, Rosen, Onosson & Li (2015) identified some significant distinctions concerning vowel quality between second-generation Filipino-Winnipeggers and their non-Filipino-ancestry counterparts. While there is certainly much more work to be done on this topic within Canada, this dissertation does not directly address ethnolinguistic differences in any respect, aside from this note.

Chapter 2 Background: North American English

This chapter presents relevant background information on North American English; it is divided into five subsections. The first two discuss the two topics of immediate concern under the research questions posed in the Introduction: the abbreviation of vowels before voiceless codas, and Canadian Raising. The three subsequent subsections provide information on the historical provenance of the English language, and the contemporary phonological and phonetic characteristics of vowels, within the three dialect regions represented in the studies carried out for this dissertation: Inland Canada (Winnipeg), The West (Denver) and The North (Madison). The historical summaries presented for each region document the periods leading up to the point at which English became the predominant spoken language, which is roughly contemporaneous with the nineteenth or twentieth centuries, depending on the region.

2.1 English pre-voiceless vowel abbreviation

In many languages, vowel durations vary systematically by way of a phonological distinction between long and short vowels which are otherwise of similar quality (i.e. articulatory position, rounding, nasality etc.), such as occurs in Japanese or Arabic. Among languages which lack such phonological vowel length distinctions, phonetic vowel duration differences are still frequently observed, falling into two categories, both of which occur in English: *differences in inherent vowel durations*, e.g. between /i/ in *heed* vs. /ɪ/ in *hid*; and *contextual differences in phonetic vowel durations* related to the voice quality of the following consonant, e.g. between /i/ in *heed* vs. /i/ in *heat*. The latter type has been well-documented not only for English (see §2.1.1 and §2.1.2 below) but also cross-linguistically across a range of languages including Hungarian (Meyer & Gombocz 1909), Italian (Metz 1914), Spanish (Navarro Tomas 1916), German (Maack 1953), Norwegian (Fintoft 1961), Swedish (Elert 1964), Danish (Fischer-Jørgensen 1964), Dutch (Slis & Cohen 1969), French, Russian, Korean (Chen 1970), Hindi (Maddieson & Gandour 1977) and Persian (Ghadessy 1986, cited in Kluender et al. 1988).[1]

Vowel duration differences related to coda voice quality, where they occur cross-linguistically, invariably show a pattern of shorter vowel durations before voiceless consonants and longer durations before voiced consonants. Although this general pattern is not restricted to English, cross-linguistic comparisons (Zimmerman & Sapon 1958; Delattre 1962; Chen 1970) indicate that it may be more pronounced in English than in some other languages. To refer specifically to the particular instantiation of this phenomenon of durational abbreviation as it occurs in English, I coin the term here *pre-voiceless vowel abbreviation*, or PVVA, leaving investigation of the relationship between PVVA and the more generally observed, cross-linguistic pattern to other research.[2] In the following two subsections, I review the literature on studies of PVVA in English, from the view of both production (§2.1.1) and perception (§2.1.2).

[1] Despite widespread occurrence, some studies on languages such as Arabic (Mitleb 1984), Czech and Polish (Keating 1985) indicate that contextual vowel duration differences may not be a completely universal property of human language.

[2] Kluender, Diehl and Wright (1988) introduce the term *vowel-length effect* or VLE to refer to the same pattern, although it is not clear whether they intend it to refer to general wider cross-linguistic pattern, or only the specific process which occurs in English; for this reason, I use different and more specific terminology.

2.1.1 PRODUCTION STUDIES OF PRE-VOICELESS VOWEL ABBREVIATION

The linguistic literature documenting PVVA in English dates at least to the beginning of the 20th century, although it has certainly been present in the English language for much longer than that[3]. Meyer (1903) investigated the speech of two individual British speakers and reported that vowels before voiced codas were 40% longer than before voiceless codas (cited in Jespersen 1954:449). Although first documented in Great Britain, this pattern is certainly not restricted to British English varieties, as it has frequently been observed that the phonetic durations of North American English vowels, too, are substantially abbreviated (to varying degrees) when preceding a voiceless consonant[4].

Systematic acoustic-based investigation of vowel duration production in North American dialects of English appears to have begun in earnest in the late 1930s and early 1940s with a series of articles in *American Speech* by Heffner and colleagues, under the heading *Notes on the Length of Vowels* (Heffner 1937, 1941, 1942; Locke & Heffner 1940; Lehmann & Heffner 1940, 1943; see also Rositzke 1939 and Heffner 1940). These studies were based on samples of the authors' (multiple) own speech and so do not substantiate any described patterns for a wider population; nevertheless, their findings set the stage for and are in broad accordance with subsequent research which has investigated PVVA more widely. In Heffner, et al.'s studies, English vowels in monosyllables containing both voiced

---

[3] Based on earlier descriptions and pairings of "long" and "short" vowels, Jespersen (1954) concludes that "This distinction seems to be at least two hundred years old," (p. 450), with the earliest such references appearing in Cooper (1685) and Elphinston (1765).

[4] While PVVA may occur in both North American and British varieties of English, it is not obvious that it has the same effect or magnitude in or throughout both regions, e.g. see Hewlett, Matthews & Scobbie (1999); the situation in other varieties is even less well-known.

and voiceless plosive codas were examined, and two general patterns were observed. First, the lax vowels [ɪ, ʊ, ʌ, ɛ] exhibit shorter durations in all contexts in comparison to the other vowels, i.e. their *inherent durations* are the shortest of all vowels. Second, all vowels are uniformly shorter before voiceless consonants than before voiced ones; the authors stress that "[t]his [durational difference, i.e. PVVA] is true for every vowel, and our evidence on this point is unequivocal," (Lehmann & Heffner 1943:212). These earliest of truly quantitative findings were corroborated by numerous later studies utilizing more sophisticated technology for audio recording and analysis, which I survey below in the form of brief summaries; where PVVA ratios are reported, these are almost always determined by my own calculations based upon reported pre-voiced and pre-voiceless vowel durations in the published articles.

House & Fairbanks (1953) investigated vowel duration in a study involving 10 speakers of "General American" (specific dialect unspecified). Vowel durations followed the PVVA pattern, with vowels before voiceless codas having a mean PVVA ratio of 0.688; that is, vowels before voiceless codas have 68.8% of the duration found before voiced codas. These differences were not only significant between homorganic coda contexts (e.g. [t] vs. [d]) but across the consonantal inventory: "All voiced environments, furthermore, produced vowels that differed significantly from all those produced in voiceless environments," (p. 108).

Peterson & Lehiste (1960) was the first major study to investigate vowel duration throughout the entire English vowel inventory. Two separate datasets were involved: a large set of 1263 words produced by one speaker, and a small set of 70 words each produced by five speakers (reported to be speakers of the same, unidentified dialect). Their main conclusion regarding the effect of coda voice quality on the preceding vowel was that "[i]n general, the syllable nucleus is shorter when followed by a voiceless consonant, and longer when followed by a voiced consonant ... the ratio of the durations of the vowels was approximately 2:3, the syllable nucleus before the voiced consonant being longer in every case," (p. 702). Averaging across all of Peterson & Lehiste's results, a mean PVVA ratio of 0.663 between coda voice contexts is obtained.

House (1961), in another study of an unspecified dialect of American English, found that duration patterns across all vowels were most significantly related to the factor of coda voice quality, and less significantly to manner of articulation. The overall PVVA ratio obtained from House's results is 0.548.

Klatt (1973) looked at the interaction of two factors, coda voice and syllable quantity, on vowel duration. Klatt elicited spoken utterances from three adult male speakers (dialect unspecified) which were randomly generated from a list consisting of monosyllabic and bisyllabic pairs with the same initial syllable, e.g. *beat* vs. *beaten, need* vs. *needle,* etc. Durational differences between coda voice contexts were significant, with a PVVA ratio of 0.667.

Umeda (1975) examined continuous speech data from three American speakers from locations representing different dialects: New York, Ohio, and "southern U.S.". Umeda's results are not aggregated in such a way to allow determination of overall PVVA ratios, but the reported vowel duration patterns consistently have shorter vowels before voiceless codas and longer vowels before voiced codas, in line with previous findings.

Several studies have looked at PVVA effects in atypical speakers; three such studies are discussed here and just below. The first of these is Sharf (1964), who looked at PVVA effects in a comparison of normal (i.e. modal phonation) and whispered speech. Three speakers of American English (dialect unspecified) produced a series of CVC syllables with varying coda voice quality, which were found to differ significantly by coda context for both types of speech. The PVVA ratios obtained from Sharf's results are 0.656 for normal speech and 0.62 for whispered speech. Sharf's finding of significant durational differences even in whispered speech, where phonation is not active, suggests that PVVA is phonological in nature, rather than deriving strictly from physiological effects from the activation of phonation in the larynx, although this of course does not rule out a historical physiology-based origin for the process.

Whitehead & Jones (1976) compared PVVA effects among three groups of speakers: normal-hearing, severely hearing-impaired, and profoundly deaf, all congenital (from birth) conditions. Ten subjects per group produced CVC syllables with varying coda voicing. PVVA ratios obtained from the reported results for each group are 0.721 for normal-hearing, 0.768 for hearing-impaired, and 0.853 for deaf speakers. ANOVA testing found that vowels in voiced and voiceless coda conditions were significantly different for all but the deaf individuals, who had the highest PVVA ratio (i.e. least abbreviation before voiceless codas). These results support the view that the PVVA effect in English is, at least in part, phonological and must be learned by exposure to spoken language, as it is most apparent in those born without hearing impairment.

Gandour, Weinberg & Rutkowski (1980) compared PVVA production results between typical speakers and those who had undergone laryngectomy (removal of the larynx) and learned to produce an approximation of phonation using the esophagus instead of the larynx. This comparison is illuminating with regard to the potential motivation underlying PVVA because, as the authors point out, "[i]f vowel-length variation induced by the voicing of the post-vocalic consonant environment in English is governed by inherent physiological characteristics of laryngeal adjustment, we would not expect to see this effect in esophageal speech due principally to the absence of normal phonatory apparatus," (p. 150). The presence of PVVA among esophageal speakers, then, would indicate that PVVA has a substantial non-phonetic motivation of some kind. Three subjects from each group, laryngeal and esophageal phonators, produced a series of CVC syllables with differing coda voicing. Both groups exhibited significant durational differences following the PVVA pattern, with esophageal speakers having longer overall vowels and larger standard deviations of vowel duration than laryngeal speakers and a lower PVVA ratio (i.e. more abbreviation) of 0.574 compared with 0.633 for laryngeal speakers.

Luce & Charles-Luce (1985) examined the factors of vowel duration, consonant duration, and the ratio between the two (the C/V ratio) under a number of test conditions. Two experiments were conducted using five male and five female subjects in total (dialect unspecified). Minimal pairs containing the vowels /i, ɪ, ɑ/ in pre-voiced and pre-voiceless context were embedded in sentence frames which positioned the target words adjacent to various consonant and vowel types. Vowel duration was determined to be the factor most consistently correlated with coda voicing; the factors of consonantal (closure) duration and the C/V ratio were also found to be correlated with coda voicing, but less consistently than duration. Collapsing together the various contexts in which the test tokens were placed, the mean PVVA ratio across all subjects was 0.69.

De Jong (1991) investigated PVVA from the view of articulatory, rather than acoustic production. Two English speakers (presumed American, dialect unspecified) were recorded via X-ray while producing a variety of alveolar-final tokens of differing voice quality. Longer vowel articulations were found to correlate significantly with the presence of voiced codas for both subjects, covering more than 25% of the observed variance in overall vowel duration across voiced and voiceless coda tokens (de Jong notes that other factors, such as consonant manner of articulation, also account for smaller proportions of vowel duration variation), providing solid articulatory evidence for the PVVA effect in English (published results do not permit calculation of PVVA ratio).

De Jong (2001) returned to an acoustic investigation of PVVA. In this study, four speakers of "midwestern" American English produced a series of nonsense tokens varying in coda voice quality (de Jong also elicited tokens of open syllables to investigate durational patterns based on onset voice quality, which I ignore here). Speakers were instructed to produce each token in time with a metronome in order to examine speech rate effects. Two series of elicitations were conducted, one with the metronome set to a fixed rate throughout, and the other with the metronome increasing in frequency from start to finish. With regard to voicing i.e. PVVA effects, for two speakers changes in speech rate did not appreciably alter the PVVA pattern; these speakers produced a very stable pattern with short vowels before voiceless codas and long vowels before voiced codas even at increasing speech rates. Additionally, while vowels before voiced codas varied in length proportional to rate, with longer variants at slow tempos, this was not the case for pre-voiceless vowels which were very stable in duration across all speech rates. However, the other two speakers did not exhibit such stability, so on the whole the results are equivocal with regard to PVVA production (durational values were not published, so PVVA ratios cannot be calculated). De Jong's results indicate that speech rate is an important factor in how PVVA is implemented, which may be overlooked in stable speech rate contexts, such as is often the case in experimental conditions, i.e. in nearly every other study of PVVA production.

De Jong (2004) further complicated the investigation of PVVA through the addition of the factor of stress placement. Five speakers of "midwestern" American English were recorded producing tokens of minimal pair syllables with differing coda voice quality (e.g. *bed* vs. *bet*) in three stress environments, with the target syllable carrying either primary stress, secondary stress, or in an unstressed position (e.g. *bed* vs. *flower bed* vs. *rabid*), as well as in two focus environments, lexically focused vs. not focused; I ignore here results pertaining strictly to durational differences between stress and focus environments which do not include coda voicing. Results indicated that the PVVA pattern was observable throughout all tested conditions, but there was a strong interaction between coda voicing with both stress and focus. Increased stress or focus were both associated with increased vocalic durational differences between coda voicing contexts, i.e. a larger PVVA effect. The PVVA effect was nearly nonexistent in unstressed syllables, and largest under primary stress; likewise, lexically focused syllables had a larger PVVA difference than non-focused syllables. The results of this study indicate, as with de Jong (2001), that PVVA is mitigated by other factors which are involved in online speech, such as stress and focus (or speech rate), that may go unnoticed under experimental conditions which do not explicitly include them.

Over the past decade or so, researchers have increasingly investigated and reported on regional differences in vowel durations (Clopper, Pisoni & de Jong 2005; Fridland, Kendall & Farrington 2013, 2014; Jacewicz & Fox 2015). However, in the vast majority of cases, vowels are elicited in a single frame with an invariant coda, e.g. *head, hide, had*, etc., which makes it impossible to report on possible PVVA differences between regions or dialects. Over the same recent period, and as of this date, I am aware of only two studies which have specifically looked at such PVVA differences across dialects: Jacewicz, Fox & Salmons (2007) and Tauberer & Evanini (2009).

Jacewicz, Fox & Salmons (2007) conducted a study involving nine female and nine male speakers from each of three U.S. dialect regions: central Ohio, south-central Wisconsin, and western North Carolina. Tokens of five minimal pairs with different coda voicing (e.g. *bites* vs. *bides*) were elicited from each speaker, allowing investigation of PVVA across five target vowels: /ɪ, ɛ, æ, e, aj/. With regard to PVVA, ANOVA testing found a significant effect of consonantal context (i.e. voicing) on vowel duration, and the authors note that "the general tendency for vowels to be longer before voiced consonants as opposed to voiceless is maintained across all vowels, all dialects, and both genders," (p. 377). However, precise values of individual vowel durations by coda context and dialect are not reported, so calculation of PVVA ratios is not possible for this study.

Tauberer & Evanini (2009) drew from the continent-wide study which contributed to the *Atlas of North American English* (ANAE; Labov, Ash & Boberg 2006). The ANAE data was force-aligned using the P2FA forced- process (Yuan & Liberman 2008), yielding vowel durations for 109,652 ANAE tokens from 514 speakers). Tauberer & Evanini report the PVVA effect as a ratio of pre-voiced duration to pre-voiceless duration; by inverting this ratio, the resulting values are comparable to the findings which have been related for the other studies summarized in this section. These inverted ratios are summarized in Table

2.1; Tauberer & Evanini include segregated results for the U.S. state of Maine and the city of Boston, which form outliers at either end of the PVVA ratio spectrum.

Table 2.1 PVVA ratios in ANAE (Tauberer & Evanini 2009)

Dialect	Ratio	Dialect	Ratio
State of Maine	0.98	North	0.813
New York City	0.885	Eastern New England	0.806
South	0.862	Southeast	0.806
Canada	0.84	Mid-Atlantic	0.8
West	0.84	Western Pennsylvania	0.787
Midland	0.826	*City of Boston*	0.752

The ratios in Table 2.1 are substantially higher than any others reported in this section for non-hearing-impaired speakers, in many cases approaching or exceeding the high ratio of 0.853 reported by Whitehead & Jones (1976) for deaf speakers. The cause for this disparity is likely due to differences in data collection methodology. Unlike the other studies in this section, which utilize word-list-based, laboratory-elicited speech, Tauberer & Evanini drew from the ANAE's corpus of sociolinguistic interview data, which has the express aim of eliciting a more casual, natural form of speech (Schilling 2013). As discussed earlier, De Jong (2004) established that the factors of stress and focus significantly affect PVVA ratios, causing larger ratios (more abbreviation) where they occur. Therefore, it seems likely that the type of careful speech which occurs in the laboratory would have the effect of exaggerating durational differences produced by PVVA, as compared to the type of speech which would be expected to occur in sociolinguistic interviews. Another possibility concerns data processing or analysis. Tauberer & Evanini utilized automated rather than manual vowel segmentation methods, which might be responsible for some of the differences observed with respect to segmental duration. Because both speech style and segmentation methods differed between Tauberer & Evanini and the other papers cited with respect to PVVA effects, determining their individual and combined effects on durational results would be a difficult task. I believe it is at least reasonable to speculate that both factors played some role in producing the dramatically different PVVA results observed.

It is also worth noting that mean vowel durations (in word-final syllables and irrespective of coda context) differ across dialects, but do not pattern in the same order as PVVA patterns; from shortest to longest mean vowel durations as calculated by Tauberer & Evanini, the major North American dialects are ordered as follows: New York City < Eastern New England < Canada < Mid-Atlantic < North < Western Pennsylvania < West < Midland < South < Southeast (compare with Table 2.1 above). The fact that PVVA differences between dialects appear to pattern differently than overall vowel duration differences highlights the need for further investigation of PVVA ratios across dialects. And, the differences in reported results between sociolinguistic interviews and laboratory-based speech indicate, perhaps paradoxically, that the "unnatural", exaggerated style elicited in a laboratory or similar setting might actually allow easier identification of such differences, by exaggerating already-present differences.

Pycha & Dahan (2016) investigated durational patterns of /aj/ before voiced and voiceless codas, using six minimal pairs e.g. *bite~bide, height~hide*, etc. embedded in a carrier phrase. Nine female speakers of a variety of American English dialects were involved in the production study. Linear mixed-effects modelling indicated that durational differences were not significantly correlated with following coda context, although it is described by the authors as having "approached significance" (β=6.46, t=1.86, p=0.06; p.

21). Taking this description with the appropriate grain of salt, the PVVA ratio determinable from their data is 0.792.

The final study which I will describe in this section is Hall (2016a,b). Although Hall's study was largely focused on comparing time-normalized durations as implemented in SSANOVAs, and not PVVA effects, her results on durational differences are especially notable because they are broken down into discrete results for male and female speakers, for each of the two Canadian cities of Toronto and Vancouver, and for each CR diphthong. As such, a portion of this data is directly comparable to the female, Winnipeg population included in this dissertation (see 3.1). In Hall's study, PVVA ratios ranged from a low 0.589 for female Torontonian /aj/ to a high of 0.717 for male Torontonian /aw/. Hall conducted linear mixed effects testing for the factors of *vowel* (i.e. one of the two CR diphthongs; no other vowels were considered), *coda context* (i.e. voiced vs. voiceless), *region* (i.e. city, Toronto vs. Vancouver), and *sex* as fixed effects, along with random intercepts for *speaker* and *word*. Unsurprisingly, Hall reports that *coda context* was significantly correlated with duration (p<0.0001); an effect of speaker *sex* was also found, albeit of weak significance (p=0.0348). The factors of *vowel (diphthong)* and *region (city)* were non-significant, indicating that variations in duration between /aj, aw/ and between Toronto and Vancouver, respectively, were non-distinctive for her speakers.

Table 2.2 PVVA ratios in Vancouver and Toronto (Hall 2016b)

	Vancouver			Toronto			
Diphthong	Female	Male	All speakers	Female	Male	All speakers	Both cities
/aj/	0.638	0.658	0.648	0.589	0.637	0.613	0.631
/aw/	0.666	0.668	0.667	0.7	0.717	0.709	0.688
Both diphthongs	0.652	0.663	0.657	0.645	0.677	0.661	0.659

Table 2.2 presents the various mean durations and PVVA ratios provided in Hall (2016b:38) as well as some means calculable from the published results, although not in the original. Despite Hall's finding that *diphthong* and *city* were non-significantly different with respect to duration when tested across her entire dataset, there are some intriguing differences between both diphthongs and across the two cities which can be observed. For example, while /aj, aw/ have very close ratios in Vancouver, at 0.648 and 0.667 respectively, in Toronto they are more disparate, at 0.613 and 0.709. Additionally, while female ratios are smaller (more PVVA) than male ratios uniformly, they are more distinctive for certain pairings. For example, for /aw/ female and male ratios are very close in both Vancouver (0.666 vs. 0.668) and Toronto (0.7 vs. 0.717), but for /aj/ they are less similar, again both in Vancouver (0.638 vs. 0.658) and Toronto (0.589 vs. 0.637).[5] Furthermore, Toronto speakers cover a wider range of PVVA ratios overall, from a low of 0.589 to a high of 0.717, while Vancouver speakers exhibit less overall variation, from a low of 0.638 to a high of 0.668, despite the two city's overall ratios being very similar at 0.661 (Vancouver) and 0.668 (Toronto). I make note of these facts not to dispute Hall's findings in any way, but rather to point out that statistical results are subject to interpretation based upon the questions posed and the ways in which they are investigated. The indication here that two major Canadian cities may vary in terms of their durational variation patterns suggests that PVVA effects within CR are deserving of more investigation within Canada.

---

[5] Hall notes the perhaps unintuitive finding that male speakers' larger ratios are the result of their smaller overall durational range; female speakers produce a wider range of durations than males, including both lengthier unabbreviated, and shorter abbreviated vowels.

2.1.2 PERCEPTION STUDIES OF PRE-VOICELESS VOWEL ABBREVIATION

While the PVVA pattern in English is well-documented in studies looking at acoustic and articulatory production as discussed in §2.1.1, its role in perception is somewhat less clear. Although many studies have found that preceding vowel duration is a significant factor in the correct identification of coda voice quality, and often the primary such factor, many researchers also argue that it is only one among a suite of features which all appear to be involved in the accurate perception of voicing, such as: *voice bar duration, consonant duration, the ratio of vowel-to-consonant duration (C/V ratio), plosive closure duration, burst/frication duration, transitional F_0 contours,* and *transitional formant frequencies* between vowel and consonant. For the purposes of this dissertation, which is focused on the production side, teasing apart these various perceptual factors is not essential. The survey of studies presented in this section is intended merely to corroborate the findings from the production studies surveyed in §2.1.1, that vowel duration differences are strongly connected to coda voicing differences, without any implication that such differences categorically determine how voicing is perceived.

One of the earliest perceptual studies related to PVVA is reported in Denes (1955). Synthesized vowels of varying durations were spliced to a recording of a naturally-spoken [s] (the equipment available to Denes at the time could not produce an authentic-sounding fricative) which was manipulated to vary in duration. 33 subjects participated in the experiment, in which they indicated their perception of each synthetic token as containing a final [s] or a [z]. Perception of a [z] (that is, of a voiced coda) was found to depend on the ratio of vowel to consonant duration (the C/V ratio); long vowels with short consonants yielded the highest rate of voicing perception, i.e. as [z], contrasting with short vowels with long consonants which yielded the lowest rate, as [s]. When the duration ratio between the two segments was approximately 1:1, perception rates for [z] were around 50%, i.e. no better than chance.

In another study of PVVA effects on perception, Raphael (1972) generated completely synthetic experimental stimuli of monosyllables mimicking voiced coda conditions, i.e. having relatively long vowels and relatively short coda consonants, covering a range of different vowel durations. A second sequence of "voiceless" stimuli were created from the "voiced" stimuli by altering the relative durations of the vowel and consonant segments. 25 participants listened to the stimuli in a forced choice experiment; for each token they heard, they selected between a minimal pair which differed only in the voicing of the coda, e.g. *bet* vs. *bed.* Raphael's findings were (nearly) unequivocal: "with one exception and regardless of the voicing cues used in their synthesis, all final consonants and clusters were perceived as voiceless when preceded by vowels of short duration and as voiced when preceded by vowels of long duration," (p. 1298).

Hogan & Rozsypal (1980) conducted a PVVA perception study which is notable in part for the fact that it was conducted in Canada, at the University of Alberta in Edmonton, and hence is likely the earliest such study to involve speakers (and listeners) of a Canadian English dialect. Stimuli for the perceptual study were elicited from a single female Canadian speaker. Analysis of voiced vs. voiceless coda ratios among the recorded stimuli obtained a PVVA ratio of 0.735, the highest among all reported ratios for non-hearing-impaired speakers (this was only a single speaker and thus should not be taken as a representative sample). The original recordings were then digitally altered by manipulating vowel duration, producing five variants per token. 14 Canadian subjects performed a forced choice task, choosing between minimal pairs differing in coda voice quality for each of the stimuli. Among the factors examined, the duration of overt voicing (i.e. a visible "voice bar") had a slightly greater effect than overall vowel duration, with these two factors covering 22% vs. 21% of the variance in participant responses, respectively (all other factors tested were below 5% each). The authors concluded that, while vowel duration is an important factor in the identification of coda voice quality, it is only one among several factors which contribute to its perception.

Wardrip-Fruin (1982) conducted a perception study using recordings of two speakers, which were then manipulated in a variety of ways including deletion or expansion of various portions of the vowel, deletion of the final consonant, and synthetic alteration of the presence of voicing during the final consonant. 12 participants listened to both the unaltered and manipulated tokens, and made a forced choice of coda voice quality, e.g. distinguishing between *bead* vs. *beat*. Wardrip-Fruin found that a variety of factors were significantly correlated with accurate identification of coda voicing, and that the combination of cues was generally more important than any single cue on its own. The absence of one cue in a particular token, e.g. coda voicing, put more weight on the remaining cues, e.g. vowel-to-coda transitional formants, in making the forced choice. For example, the presence vs. absence of the coda segment itself had a greater effect on accurate identification of voicing than any aspect or manipulation of vowel duration, but when the final segment's acoustic information was deleted total syllable duration was more significant than mere vowel duration.

Soli (1982) conducted a series of experiments looking at internal dynamics of vowels in pre-voiceless and pre-voiced contexts to investigate whether factors beyond overall vowel duration were related to perception of coda voicing, based on the hypothesis that "modifications in vowel duration are achieved by a temporal reorganization of the entire syllabic gesture which alters the dynamic formant structure of the vowel," (p, 367). Four related experiments were conducted involving synthetically altered tokens by varying vowel and consonant durations and adjusting the internal vocalic spectral structure while maintaining duration constancy. In particular, the portion of the vowel composed of the initial steady state phase was a target of manipulation. Subjects performed a discrimination task to identify each token as either the noun *(the) use* or the verb *(to) use*. Results indicated that vowel duration was by far the most significant correlate of accurate token, i.e. coda voice, identification.

Port & Dalby (1982) performed a set of experiments to investigate perception of coda voicing, focusing on the ratio of coda consonant-to-vowel duration (C/V ratio), rather than the absolute duration values of either segment alone, as a potentially important cue for coda voicing. Synthetic stimuli were created and manipulated to produce a range of different vowel and coda consonant durations. Regression testing over three related experiments indicated that the factor of vowel duration alone had a larger correlation (R^2 = 0.629; 0.698; 0.619) with correct identification of coda voicing than either the factor of coda duration (R^2 = 0.564; 0.475; 0.526) or the C/V ratio (R^2 = 0.610; 0.661; 0.578), although it should be noted (and the authors argue that) the C/V ratio was found to be nearly as explanatory as vowel duration, in each experiment[6].

[6] Port & Dalby (1982) also argue that the C/V ratio is potentially more useful in perception as it is more stable across varying speech rates; as this dissertation is not focused on speech perception, I will not discuss this point further, but see Massaro & Cohen (1983) for a direct and immediate counterpoint response to Port & Dalby.

On the whole, acoustic perception studies investigating PVVA phenomena are somewhat more equivocal than production studies in determining the function of vowel duration, commonly noting other factors which are also significantly involved in the perception of coda voicing. Nonetheless, the results forthcoming from the perception studies surveyed in this section strongly affirm the role that relative vowel duration plays in signalling the voice quality of the following coda segment. Taken together with the findings on the production side, they indicate that PVVA is a robust part of North American English phonology across a variety of dialects. That being said, differentiation of PVVA effects between and across dialects is not very well studied or understood at this time.

2.2 Canadian Raising

The term Canadian Raising (CR), as noted in the Introduction, refers to allophonic variation of the diphthongs /aw/ and /aj/ among speakers of Canadian English, for whom it has become something of a hallmark and a stereotype, especially among Americans, who sometimes speak of Canadians saying *oot and aboot* (Boberg 2008; Sadlier-Brown 2012; Menclik 2013). While there is some debate as to the history and origins of CR, its occurrence in Canada has been documented since at least the 1930s.

One of the earliest references to diphthong-raising in Canada is a short passage relating the raising of /aj/ and /aw/ in Ontario, within a study otherwise focused entirely on local speech in Virginia (Greet 1931). Perhaps the earliest works focused specifically on Canada which reference diphthong-raising are Ahrend (1934) and Ayearst (1939), both of whom refer explicitly only to raising of /aw/: “raising, among Canadians, especially of the diphthong [aʊ],” (Ahrend 1934:136); “the Canadian appears to say [u:t]. Actually he says [aʊt] but the word is clipped,” (Ayearst 1939:231–2). The first Canadian-focused and truly phonological account of raising of both /aj/ and /aw/ is found in Joos (1942), who describes the process as follows:

> “the diphthongs /aj/ and /aw/ … each have two varieties. One, which I shall call the HIGH diphthong after its initial tongue-position, begins with a lower-mid vowel-sound; it is used before any fortis consonant with zero juncture … The other, the LOW diphthong, is used in all other contexts,” (p. 141).

Joos’ analysis of diphthong-raising in Canadian English was followed most famously by Chambers (1973, 1975, 1981, 1989, 2006), who appears to have coined the term *Canadian Raising* and was the first to provide descriptions of it (in the framework of generative phonology), as in Figure 2.1.

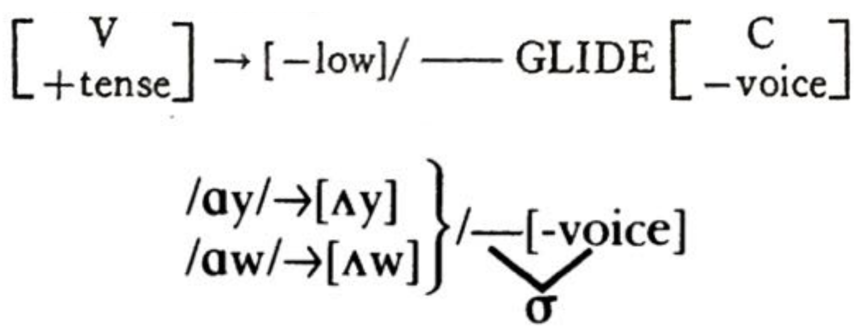


Figure 2.1 Two versions of a generative-phonology, feature-based rule for Canadian Raising (Chambers 1973:116, 1989:79)

CR has been discussed under a variety of phonological theories and models including generative phonology (Chambers 1973, inter alia; Picard 1977), autosegmental phonology (Paradis 1980), optimality theory (Myers 1997; Pater 2014), and exemplar theory (Hall 2007), etc. Although taking disparate theoretical approaches to CR, these are all in broad agreement with respect to its form, i.e. the occurrence of raised diphthong nuclei in allophones of /aj, aw/, and its environment, i.e. before voiceless codas. In other words, the base description of CR in the Canadian context has not changed substantially over more than seven decades of research[7].

As mentioned above, the historical provenance of Canadian Raising is a matter of some debate. Gregg (1973) suggests a line of development having the "raised" form /əi/ originating first, deriving from Middle English /ī/ during the Great Vowel Shift, with the form /aɪ/ (Gregg's notation) occurring over time in an increasingly broad set of postvocalic environments, as in Figure 2.2.

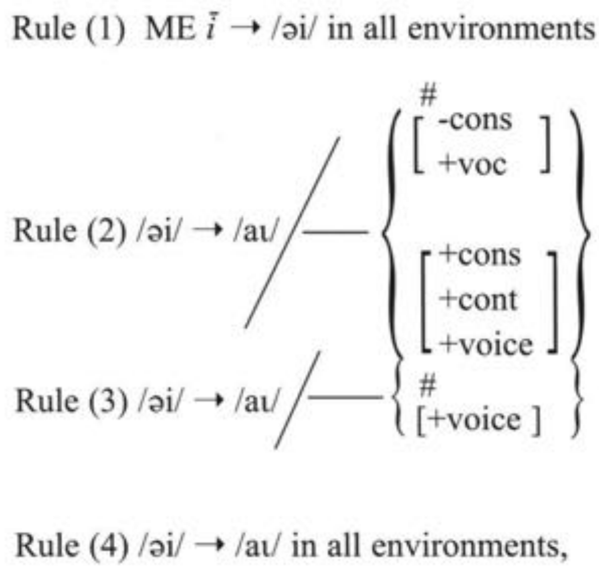


Figure 2.2 Historical development of the diphthong /aj/ (Gregg 1973:240)

Gregg's Rule (3) denotes the origin of CR itself in Canadian English; Rule (4) represents both contemporary Standard British and American English, i.e. completely non-raising

[7] There is a fair amount of discussion as to the interaction of CR with features such as stress patterns (e.g. Chambers 1989; Myers 1997; Bermúdez-Otero 2014), intervocalic flapping (Chambers 1973; Myers 1997), lexical neighbourhood effects (Hall 2005a,b), etc.; I do not focus on any of these areas in this dissertation.

dialects wherein lowering of /əi/ to /aɪ/ occurs wholesale. While the particulars of Gregg's account are not of immediate concern for this dissertation, it ends by listing a series of "relevant matters" (not elaborated on), including among them "the promising possibility of tying in the diphthongs *əi* and *aɪ* with a feature [±length]," (p. 240). Here Gregg references a potential connection between CR and PVVA. As noted in §2.1 PVVA, like CR, has a centuries-old history in English (see Footnote 3). A historical relationship between the two phenomena was claimed earlier by Joos (1942), who describes the phonologization of CR as "a *shift from a difference essentially of length* to a difference essentially of quality, so that in /aj, aw/ the difference between pre-fortis [i.e. pre-voiceless] and other articulation *is not the same* as it is for all other syllabics (including /oj/)," (p. 142, emphasis added). Chambers (1973), too, comments on the historical connection between PVVA—which he terms simply *Shortening*—and CR. Chambers suggests that this occurred because the low-rising diphthongs /aj, aw/ have the furthest articulatory distance to travel, and that reducing the temporal length of the vowel reduces the capacity for the speaker to fully produce the entire diphthong's articulation, from low nucleus to high off-glide. One solution to this articulatory dilemma would be to lower the off-glide target, i.e. diphthong "flattening"[8], and the other is to raise the nucleus, i.e. CR; "[f]rom this perspective, Shortening is seen as the historical precursor of Raising, or, put another way, Raising arises as a reflex of the Shortening rule," (p. 119).

[8] It is worth noting that although this possible solution suggests that dialects with flattening would favour it in abbreviated contexts, i.e. before voiceless codas, a fairly recent description of diphthong-flattening as it occurs in the American South (Moreton & Thomas, 2007), under the moniker Southern Glide Weakening, suggests that it is *not* an example of this, as "the *more-diphthongal* allophone occurs in the short voiceless environment," (p. 7, emphasis added).

Myers (1997) raised an insightful question on the effect of Chamber's Shortening rule (i.e. PVVA) on CR, namely whether it was categorical in nature. In other words, were raised CR diphthongs shortened or abbreviated in a consistent manner, as might be expected if there were some autosegmental-type feature being altered in the phonology? To answer this question, Myers conducted a small study of three female Torontonians who read from a wordlist of diphthongal monosyllabic words, including minimal pairs in terms of coda voicing and hence whether CR applies, e.g. *price, prize*. Myers found that CR occurrence itself was categorical, as expected; measurements of F1 (taken at 1/3 vowel duration) were not significantly different between CR-context diphthongs despite differences in coda manner, e.g. between *stripe* (plosive coda) and *strife* (fricative coda). However, vowel duration patterns within CR contexts turned out be non-categorical. Raised diphthongs were on the whole shorter than non-raised diphthongs, but there were also differences within each group; for example, final plosives produced significantly shorter vowels than final fricatives, both for raised and non-raised forms, e.g. the vowels of *pride* and *stripe* were significantly shorter than those of *prize* and *strife*, respectively (cf. Peterson and Lehiste 1960, who reported this pattern throughout the vowel system of American English). Although Myers' study was small, I think it makes an important point. While PVVA and CR may indeed be intimately connected, they are also distinct phenomena, both historically and contemporaneously. PVVA in effect sets the stage for CR, but does not precisely determine its form (cf. Chambers' "off-glide flattening" alternative), and this is true whether viewed diachronically or synchronically.

Despite the broad agreement on the basic description of CR noted at the beginning of this section, the characterization of the two *prima facie* distinct phenomena of /aj/-

raising and /aw/-raising as a singular phonological process has been questioned over time, notably by some prominent researchers on the topic. In his original article on CR, Chambers stated that "[t]he appropriateness of the term [Canadian Raising] resides in the *relative role* the rule plays in Canadian English, where its effect is the most readily identifiable trait of the dialect" (Chambers 1973:113, emphasis added). Some years later, Chambers also commented:

> "The fact that /aw/-raising and /ay/-raising exist independently of one another in the phonologies of regional accents *calls into question my original analysis* [i.e. Chambers 1973] of them as a single process under the rubric 'Canadian Raising' … As we discover constraints that apply to one but not the other and phonetic changes that affect one but not the other, the term 'Canadian Raising' seems appropriate only as a dialectological reflex for the *coexistence* of the two very similar allophonic reflexes in the same accent and less appropriate as a *theoretical phonological term for a single process* that affects two different nuclei," (Chambers 1989:77, emphasis added).

Boberg (2008) examined the relationship between the degree of raising of /aj/ and /aw/ separately, among 86 Canadian subjects and found "virtually no correlation … suggesting that those speakers who raise /aw/ the most do not necessarily also raise /ay/ the most and that *these vowels should be analyzed separately*," (p. 139, emphasis added). Perhaps, then, CR is best understood as a term for a dialectal feature of Canadian English, in the same way that an isogloss boundary may identify the co-occurrence of multiple distinct yet distinctive lexical forms, rather than as a single phonological *rule* (however understood) present in the grammar of Canadian English. Although a number of other researchers have noted specific differences in the phonetic realizations and behaviour of each of the CR diphthongs (Boberg 2008; Sadlier-Brown 2012; Pappas & Jeffrey 2014), the implication in Chambers' and Boberg's statements that CR *is not necessarily a single phenomenon* appears not to have caught on widely, at least within the domain of research on Canadian English.

Whether or not CR is a unified phenomenon has some implications for its association with Canadian English, something which has been at times overstated (although this seems to be less true over time). For example, Householder (1983) claimed that the isogloss boundaries for CR forms have a "peculiar coincidence" with the Canadian-American border "all the way from the Pacific to the Atlantic" (p. 7), a statement which cannot be taken as literally true. Not only is CR not universally present from coast to coast in Canada (see the isogloss boundary for Canadian Raising in Figure 2.7 in §2.3.1), neither is it exclusively Canadian; the occurrence of CR-like diphthong-raising patterns for /aj/ or /aw/[9] alone, or in tandem, has been well-documented in many English-speaking regions outside of Canada. For example, in Greet's (1931) examination of phonographic recordings of speech made in Williamsburg, Virginia, he documented a CR-like pattern for /aj/, even using the term *raising*, albeit not with reference to the vocalic nucleus:

> "[aɪ] tends toward [ɑ] or [a] in I, *mind, find, my, why, while*, and *by it* but of course the sound seldom entirely loses its diphthongal character … An

[9] Descriptions of dialects with /aw/-raising which is not also accompanied by /aj/-raising are exceedingly rare; there is but a single extra-Canadian example in Table 2.3 below, although recall that in their early descriptions of Canadian English, Ahrend (1934) and Ayearst (1939) describe /aw/-raising, but make no mention of /aj/-raising.

> opposite tendency often appears in *like, night, quite* and *right*. The first element [i.e. the nucleus] is shortened and the second [i.e. the off-glide] seems to be raised and to have acquired finally a certain consonantal character," (p. 166)[10].

Many studies across a variety of regions and dialects of English have related similar findings, wherein diphthongs were found to be produced with raised variants before voiceless consonants (although not necessarily exclusively or universally)[11], as noted in the list in Table 2.3 (not intended to be an exhaustive listing); in some studies, as indicated with an asterisk, raised diphthongs were also noted to be shorter in duration than their non-raised counterparts, harkening back to the aforementioned connection with PVVA.

[10] It is worth noting that this early description of diphthong-raising differs from most subsequent descriptions, in that raising is not characterized as applying to the nucleus.

[11] For example, in his description of speech in Martha's Vineyard, Massachusetts, Labov (1963) arrived at the following "consonant series from most favoring to least favorable to centralization [i.e. raising] … /t, s: p, f: d, v, z: k, θ, ð: 1, r: n: m/" (p. 290), which proceeds generally but not systematically from voiceless to voiced, and is described as not applying categorically to all potentially applicable lexical items within that dialect.

Table 2.3 Reported occurrence of extra-Canadian CR-like diphthong height alterations

Location(s)	Raised diphthong(s)	Study
Eastern Virginia	/aj, aw/	Shewmake (1925, 1943), Tresidder (1941)
Virginia (Williamsburg)	/aj, aw/*	Greet (1931)
Virginia	/aw/	Tresidder (1943)
Massachusetts (Martha's Vineyard)	/aj, aw/	Labov (1963)
Virginia, Maryland & N. Carolina; coastal S. Carolina, Georgia & Florida	/aj, aw/*	Kurath & McDavid (1965)
Scotland†; N. Ireland†	/aj/	Gregg (1973)
Illinois (Chicago)	/aj, aw/	Kilbury (1983)
St. Helena†; Bahamas; Tristan da Cunha†; Bermuda	/aj, aw/	Trudgill (1986)
Saba (Netherlands Antilles)	/aw/	Trudgill (1986)
Minnesota (Minneapolis); New York (Rochester)	/aj/	Vance (1987)
Ohio (Columbus), Michigan, western Pennsylvania; Minnesota	/aj/	Thomas (1989)
Iowa, Nebraska	/aj/	Allen (1989)
Minnesota, North Dakota	/aj, aw/	Allen (1989)
Ohio	/aj/	Thomas (1995)
England (Newcastle upon Tyne)†	/aj/*	Milroy (1996)
Michigan (Ann Arbor)	/aj, aw/	Dailey-O'Cain (1997)
England (Eastern Fens)†	/aj/	Britain (1997)
Vermont	/aj/	Roberts (2007)
Ohio (Cleveland)	/aj/	Moreton & Thomas (2007)
Pennsylvania (Philadelphia)	/aj/	Fruehwald (2008)
Pennsylvania (Philadelphia)	/aj/*	Fruehwald (2013)

†Non-North American location
**Raising associated with shortening*

As noted in Table 2.3, patterns of abbreviated vowel durations—often described as "short", "fast", or "quick"—occurring before voiceless consonants (i.e. PVVA) have been observed in combination with diphthong-raising in several studies which document raising in extra-Canadian contexts, including parts of the Atlantic coast of the United States and England, both regions geographically discontinuous with General Canadian English. Although the majority of studies on diphthong-raising do not describe concomitant shortening, it should be emphasized that absence of evidence is not evidence of absence; it may be that this aspect of the phonetic character of raising was either not noticed or simply not commented on by researchers. For example, Fruehwald (2013) reports durational differences between pre-voiced, unraised and pre-voiceless, raised variants of /aj/ although this is not discussed in the earlier Fruehwald (2008), for the same speech community.

Despite the recognition of the connection between PVVA and CR (e.g. by Chambers, Gregg, and Myers), and the association of raised forms with short vowels noted for some studies in Table 2.3, the function of vowel abbreviation within contemporary CR production remains somewhat obscure. The only study to date which reports data on vowel abbreviation patterns (i.e. PVVA) in the context of "full CR"—that is, inclusive of both /aj/- and /aw/-raising—appears to be that of Hall (2016a,b). Hall elicited tokens largely concentrated on the CR diphthongs (with several reference vowels also elicited) from 15 female and 15 male speakers each from Toronto and Vancouver. Hall found that /aj/ and /aw/ both consistently exhibit durational abbreviation in pre-voiceless (i.e. raising) contexts, with /aj/ exhibiting greater abbreviation than /aw/ in every location/gender demographic category, with PVVA ratios (converted from Hall's reported % of vowel reduction in raised forms) ranging from 0.717 (male Torontonian /aw/) to 0.589 (female Vancouverite /aj/), and an overall mean ratio of 0.656 (p. 38).

Hall's study is notable and exceptional in that it incorporates duration and formant values together in an analysis of CR using the smoothing spline analysis of variance or SSANOVA, a technique which is also utilized in this dissertation (§4.4.2); to my knowledge, Hall's is the first such application of SSANOVA techniques to CR data specifically. However, as Hall acknowledges, her implementation of SSANOVA does not fully incorporate the observed durational differences between coda voicing contexts: "the methods used here normalize vowel duration across tokens, and therefore had to be supplemented by separate analyses of duration … [which] do not provide any information about the rate of trajectory change over the course of diphthong," (pp. 46–47). Hall suggests (p. 47) the use of spectral rate of change (Fox & Jacewicz 2009) as one means of achieving a more satisfying and holistic comparison. While Hall's time-normalized SSANOVA method is probably the most common implementation of that technique, an alternative method for incorporating information on contextual durational differences (PVVA) directly within SSANOVA analysis is presented within this dissertation, as described in §4.4.

2.3 English in Canada

In this section, the history and contemporary characteristics of Canadian English, and specifically the dialect spoken in Winnipeg, Manitoba, are described. The primary sources consulted for this section include Labov et al. (2006) and Boberg (2010) for Canadian English, the latter especially with respect to historical development, Friesen (1987) for the general history of Manitoba, and Hagiwara (2006) for the phonetics of contemporary Winnipeg English.

2.3.1 A brief linguistic history of Canada and Manitoba

As the title of this subsection indicates, a cursorial overview is given here of the linguistic history of the territory of modern Canada, focusing on the province of Manitoba and the establishment of its English-speaking population. Readers interested in more in-depth coverage of the history of Canadian English are directed primarily to Boberg (2010), which is the most recent and probably the most definitive work of this nature, and certainly the most suited to linguistic researchers. There is, as far as I am aware, no such general work describing the history of Manitoba English specifically.

Figure 2.3 Geo-political map of Canada

The map in Figure 2.3 illustrates the current political boundaries of and within Canada. The earliest European exploration and settlement of the territory comprising modern-day Canada was largely conducted by the French, occurring at first mainly along the St. Lawrence River, including the cities of Québec and Montréal, and then further west to the Great Lakes and the interior parts of the continent. British expeditions in the late sixteenth and early seventeenth centuries later sailed into and explored parts of what is now northern Canada and established fur-trading posts under the mandate of the Hudson's Bay

Company. The westernmost of these early posts was established at the now-defunct York Factory, somewhat to the south of present-day Churchill, Manitoba along the shores of Hudson Bay. Through the 16^{th} to 18^{th} centuries, British settlers in the continent of North America, the majority situated in the colonies that would later become the United States of America, rapidly outnumbered French-origin settlers, including populations in Acadia (which included Canada's later Maritime provinces and part of the U.S. state of Maine) and Québec. Military conflicts between France and other European powers in North America and elsewhere would eventually lead to the loss or sale of most of France's colonial possessions. In 1763, the Treaty of Paris transferred possession of all of France's Canadian territory to the British, although substantial numbers of French speakers remained in what would become Canada (as well as regions of the United States) and their French-speaking descendants still reside in Québec, as well as the other provinces and territories[12].

Following the American Revolutionary War and subsequent Declaration of Independence from Britain in 1776, an influx of Americans loyal to the British Empire (United Empire Loyalists) migrated into Canada, which remained a British colony at the time. These English-speaking Americans would serve as an important input source for what would become the Canadian English dialect spoken in Ontario, eventually spreading farther west. Along with the British settlers already present in Canada, the Loyalists were

[12] The majority of French speakers outside of Québec are found in the provinces of Ontario (493,295) and New Brunswick (233,530); the other provinces and territories, including Manitoba, claim fewer than 100,000 speakers each, most well below that figure (Statistics Canada, 2011a).

joined by further migration from Britain, including Irish and Scottish migrants (with varying degrees of English fluency) who also contributed to the developing dialect.

English-speaking Scots formed an especially important group in the early history of Manitoba. In 1811, Thomas Douglas, the 5th Earl of Selkirk was granted a land concession from the Hudson's Bay Company, comprising most of present-day southern Manitoba, for the establishment of a colony, as illustrated in Figure 2.4.

Figure 2.4 The Selkirk Concession: *The Red River Colony*, or *Assiniboia*, 1817

The main colony site was located at Fort Garry (circled in red), near the confluence of the Assiniboine and Red rivers. The remains of Upper Fort Garry (distinguished from Lower Fort Garry, established further north along the Red River towards its mouth at the south end of Lake Winnipeg) are located near an area known today as The Forks, at the core of present-day downtown Winnipeg. The concession covered areas which today form parts of both the United States and Canada; the line in Figure 2.4 indicating 49° latitude, extending westwards from the Lake of (the) Woods, demarcates the present day border between Manitoba (Canada) to the north, and North Dakota and Minnesota (United States) to the south.

Three separate waves of colonists were recruited by Douglas from Scotland and Ireland (mostly the former) between 1811–1815, numbering less than three hundred in total over that period. The Selkirk Colony was not very successful during these early years due to poor weather and increasing hostility from the Hudson Bay Company's main rival in the fur trade, the North West Company and affiliated Métis populations in the area. Hostilities increased and eventually culminated in the Battle of Seven Oaks, which saw Selkirk's colonists driven from the settlement. Lord Selkirk eventually secured peace, and in 1817 signed a treaty with five "Chiefs and warriors of the Chippeway [Ojibwe] or Saulteaux Nation and of the Killistine or Cree Nation," (Oliver 1915:1288–1289), allowing the formal return of the colonists.

The Selkirk (or Selkirk-Peguis, after Chief Peguis of the Ojibwe/Saulteaux, one of the co-signatories) Treaty paved the way for the continuation and expansion of settlement of non-Indigenous people within the territory of present-day Manitoba throughout the mid-19th century. In 1867, several of Britain's North American colonies united under the British

North American Act to form the nation of Canada, initially consisting of the provinces of Ontario, Québec, New Brunswick and Nova Scotia; Manitoba soon thereafter joined the new confederation as its fifth province, in 1870. This prompted a series of formal treaties to be made between the Canadian government and Manitoba's Indigenous populations, as the Selkirk Treaty had been a private, non-governmental treaty made between Lord Selkirk himself (now deceased) and the signatory Chiefs.

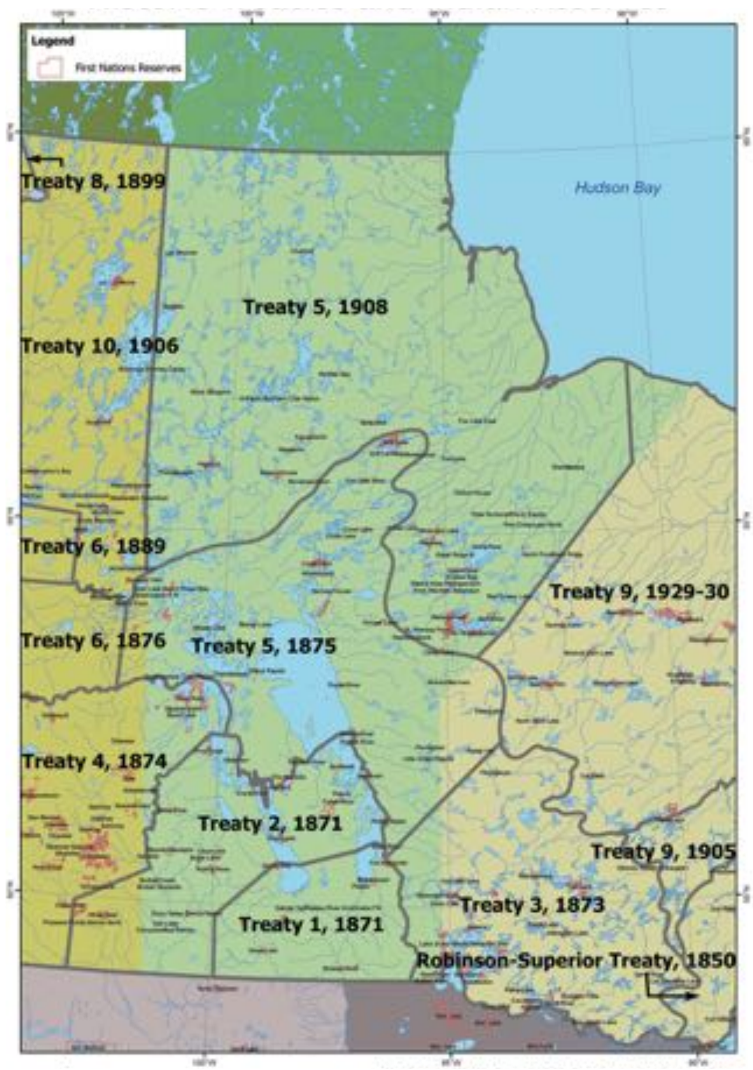


Figure 2.5 Historic Treaties and Indian Reserves in Manitoba
© Adam Downing, Manitoba Wildlands

Figure 2.5 displays the territories covered by various treaties made in the 19th and 20th centuries between the Canadian government and Indigenous groups in Manitoba Indigenous groups within the present boundaries of Manitoba when the treaties were signed included the Cree, primarily located in the north including the region around York Factory but also present throughout the province, the Ojibwe (also: Ojibway, Anishinaabe, Saulteaux) mostly to the south, the Oji-Cree in eastern Manitoba, as well as several smaller First Nations including the Dakota, Assiniboine, Dene, and also the Inuit and Métis. All of these groups have descendants who make up a substantial portion of the modern-day population of the province, and today Manitoba has the highest proportion of Indigenous ancestry of any Canadian province at 15.5%[13]. Treaty 1 includes the territory of the Ojibwe and Cree First Nations and the Métis (although the latter were not part of the treaty) and the city of Winnipeg, itself home to well over half of the present-day population of Manitoba (Statistics Canada 2016, 2017).

After Manitoba's joining of confederation in 1870, the colony of British Columbia joined with Canada the following year in 1871, and by 1905 all of the modern provinces of Canada had entered into confederation, with the exception of Newfoundland and Labrador, which had been settled in the 16th century but governed since that time as an independent colony, only joining Canada in 1949. Settlement of western Canada expanded greatly during the early 20th century, partially with the aim of forestalling American expansionary interests in the region.

---

[13] By proportion of total population, the next three provinces with the highest rates of Indigenous ancestry (excluding the federal Territories) are Saskatchewan at 14.8%, Alberta at 5.8%, and British Columbia at 4.8%; all other provinces have substantially smaller proportions, with Indigenous people making up 2% or less of the population (Statistics Canada 2006).

The perceived historical uniformity of Canadian English outside of the Atlantic region has often been attributed to the relatively rapid expansion of English speakers from Ontario westwards during this early period of confederation, in the late 19th and early 20th centuries, although there is some disagreement about the relative importance of the American vs. British dialects spoken by migrants to Canada (see e.g. Ayearst 1939; Chambers 1973; Labov et al. 2006; Boberg 2010; Dollinger & Clarke 2012). Within the then-borders of Manitoba[14], an 1870 census reported a population of 12,000 (excluding First Nations) comprised of about half French-speaking Métis, one-third English-speaking Métis, and less than one-sixth European-origin, immigrants or Canadian-born. By 1886 a huge influx of migrants from eastern Canada and immigrants from Britain swelled the population to over 100,000, 24% of English-origin, 24% Scottish-origin, and 20% Irish-origin, overwhelming and ousting the Métis (both English- and French-speaking) from their formerly solid majority position to only 7% of the population (Friesen 1987:201–202), and cementing the linguistic character of Manitoba as overwhelmingly English, which it has remained; currently, just under 73% of Manitobans report being monolingual in English (Statistics Canada 2011b).

Relative to other Indigenous North American languages, Manitoba has substantial populations of Cree (20,000), Ojibwe (9,000), and Oji-Cree (7,000) speakers, as well as a Dene-speaking community (1,000); other Indigenous languages in the province number well under 1,000 speakers. Sizeable communities of non-indigenous, non-English languages include German (70,000), French (48,000), Tagalog (40,000), Ukrainian

[14] At confederation and for the subsequent decade, Manitoba comprised only a tiny portion of its current territory, barely 5% of its present area, leading to its nickname as “the postage-stamp province”.

(19,000) and Punjabi/Panjabi (11,000). Within Canada as a whole, only a handful of Indigenous languages have a non-threatened status, which include all four Indigenous Manitoba languages mentioned by name above, although nationally both Cree and Ojibwe have somewhat declined in use recently, whereas Oji-Cree and Dene have increased slightly in usage. At present, just under one million people of Aboriginal descent (Statistics Canada's terminology) report the ability to understand any Aboriginal language, the majority not with a high degree of fluency (Statistics Canada 2012). Little to no research has been done on the possible influence, if any, that Indigenous languages or Indigenous English dialects may have had on the phonology of non-Indigenous Canadian English dialects, including Standard or General Canadian English (see §2.3.2). However, given the large proportion of population with Indigenous ancestry in Manitoba, both prior to and after confederation, continuing until the present day, if there are any such influences Manitoba would be a prime location among the Canadian provinces in which to investigate them (in the northern Territories, this proportion is of course much higher).

2.3.2 THE VOWELS OF CANADIAN ENGLISH

Boberg (2010) introduces *Standard Canadian English* as a counterpart to *Standard American English*, intended to be a broad term which covers the "geographic range [which] hypothetically extends from Victoria, British Columbia, in the west to Halifax, Nova Scotia, in the east," (p. 107). Boberg's use of this term is not intended to indicate that substantial differences do not exist within this space. The distinctiveness of the speech of Maritime Canadians, owing to that region's longer and distinct history of settlement, is widely known among lay speakers within the country as a whole, and the even greater distinctiveness of Newfoundland speech is immediately evident from its exclusion in Boberg's definition.

Current research in Canadian sociolinguistics is concerned with both identifying the unique characteristics of Canadian English which distinguish it from other global English varieties while also teasing apart its internal differences; CR is a widely-accepted example of the former (§2.2). An example of the latter is *Canadian Shift* (Clarke, Elms & Youssef 1995; see also Boberg 2005; Sadlier-Brown & Tamminga 2008; Roeder & Jarmasz 2010; inter alia), whose description has helped to identify "geographic distinctions between an 'Inland Canada' region centered on the Prairie Provinces, and areas with more variable patterns, including the larger metropolitan areas of Vancouver and Toronto," (Labov et al. 2006:217).

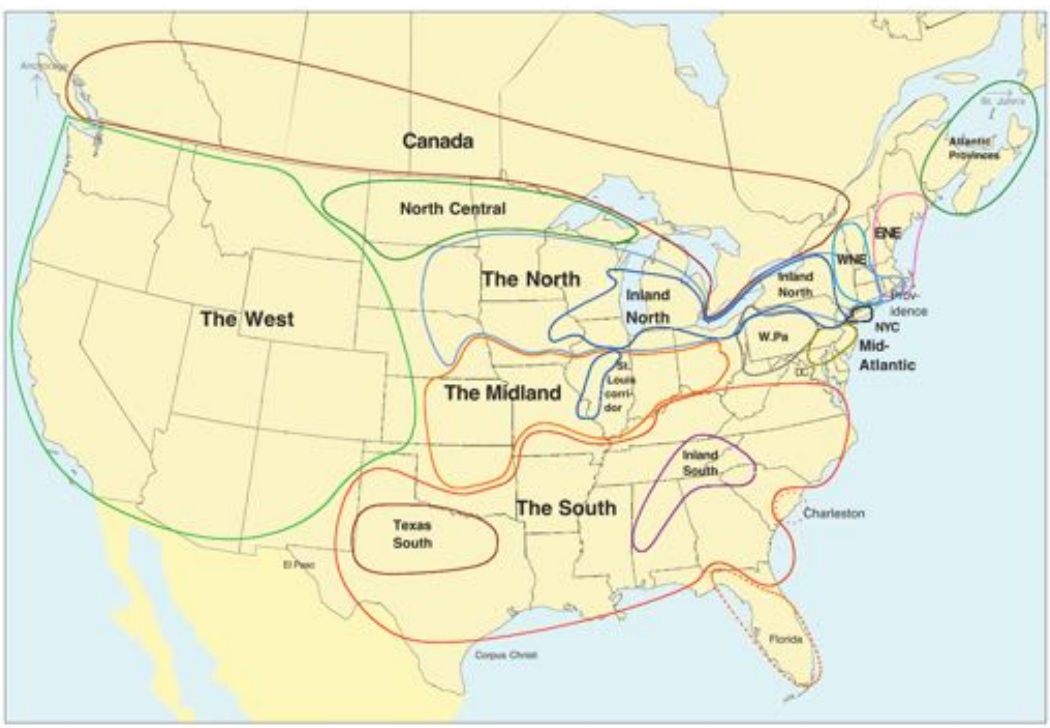


Figure 2.6 An overall view of North American Dialects (Labov et al. 2006:148)

Figure 2.6 illustrates the subdivision in Standard Canadian English between the Atlantic Provinces and what is simply termed 'Canada'. Use of the descriptor *(General) Canadian English* henceforth refers to the dialect(s) spoken in this latter region, and may contrast with the broader *Standard Canadian English*, or more localized variants such as *Atlantic Canadian English*, etc.

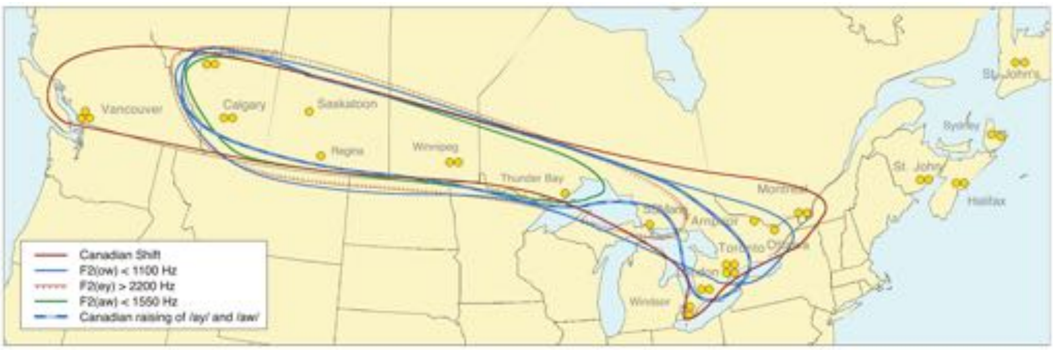


Figure 2.7 Inland Canada (Labov et al. 2006:224)

Figure 2.7 illustrates the Canadian English internal sub-division of Inland Canada, bound by the bundle of largely overlapping isoglosses that indicate the defining criteria for this sub-dialect. As noted in Figure 2.7's legend, several of these features centre around F2 values, but Canadian Shift and CR (as described by Labov et al.) in particular are not determined exclusively or primarily from F2. Inland Canada covers the region extending roughly from Edmonton in the northwest, to Toronto in the southeast. Boberg (2008, 2010) proposes further sub-divisions within Inland Canada than those identified by Labov et al., but which are already visible in the isogloss patterns seen in Figure 2.7. Boberg's major proposed sub-division is between the West and Southern Ontario (e.g. Toronto), with cities in Northwestern Ontario such as Thunder Bay falling into "a wide and sparsely populated transition zone" (Boberg 2008:152) between the two[15]. A minor sub-division is also proposed which separates British Columbia from the remainder of The West; the city of Winnipeg falls close to the centre of the latter region and is arguably the main historical centre for the dialect due to its early settlement relative to the other regional population centres. As this dissertation is not principally concerned with investigating internal Canadian sub-dialectal divisions, discussion will now turn to the characteristics of the General Canadian English vowel system.

As with many other dialects of English, the primary features which are typically argued to distinguish Canadian dialects from other varieties concern its vowels. The phonology of Canadian English is in many respects similar to that of several American English dialects. Leaving aside particular phonetic realizations, the distribution of vowel phonemes is largely the same as other varieties which exhibit the low back vowel merger, e.g. of *cot* and *caught*; non-merging dialects have an additional vowel phoneme in this part of the vowel space, but the distribution of vowels is otherwise quite similar if not identical.

[15] This author can speak from personal experience that this is an apt characterization, having made the trip between Winnipeg and Toronto by car on several dozen occasions; much of the multi-thousand-kilometre distance is linked only by an undivided, two-lane highway. In 2016, a bridge failure along this route resulted in the severing of "the only road connecting Eastern and Western Canada," (Husser 2016).

Table 2.4 The vowels of Canadian English, adapted from Labov et al. (2006:11-12)

	SHORT		LONG			
	V		Front upgliding Vy		Back upgliding Vw	
nucleus	front	Back	Front	back	Front	back
high	/i/ [ɪ] *bit*	/u/ [ʊ] *put*	/iy/ [ij] *beat*		/iw/ [ɨw] *suit*	/uw/ *boot*
mid	/e/ [ɛ] *bet*	/ʌ/ *but*	/ey/ [ej] *bait*	/oy/ [ɔj] *boy*		/ow/ *boat*
low	/æ/ *bat*	/o/ [ɑ] *cot, balm, caught*		/ay/ [aj] *bite*		/aw/ *bout*

The transcription of vowel phoneme inventory of Canadian English in Table 2.4 follows the format used in Labov et al. (2006) and is adapted from their Tables 2.1, 2.2 and 2.3 (pp. 11-12). The phonemic transcriptions used are based on a binary system which focuses on identification of lax and tense ('SHORT' and 'LONG' in Table 2.4) vowel counterparts through use of the same primary symbol, with the members of each counterpart pair distinguished by the presence or absence of an off-glide. The phonetic transcriptions indicated in Table 2.4 are drawn from those used in Kurath (1977) and added only where the phonemic transcription differs from standard IPA notation; italicized keywords used by Labov et al. also accompany each phoneme, for clarity.

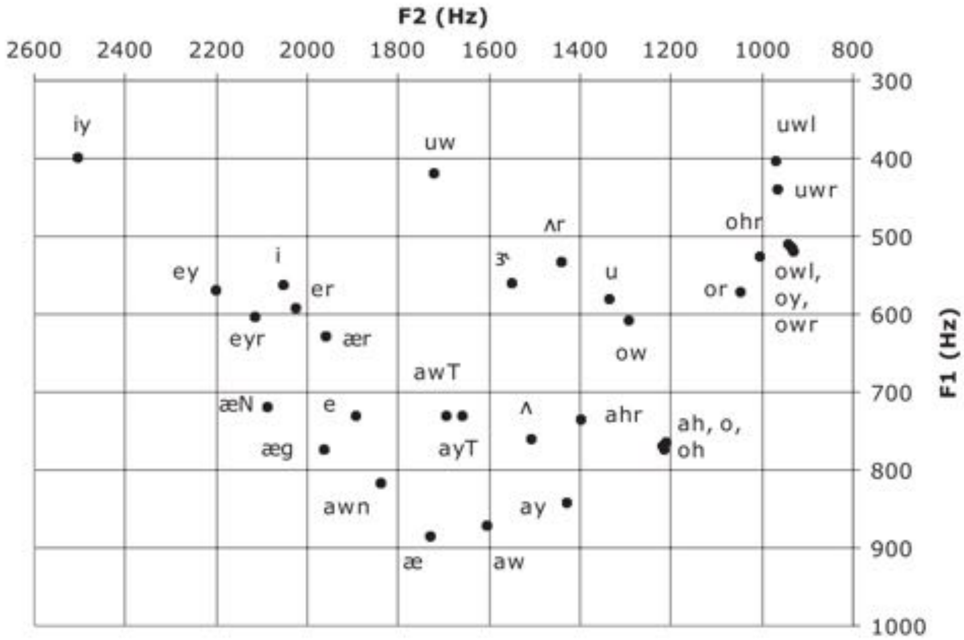


Figure 2.8 Mean F1 and F2 Measurements for Vowel Phonemes and Major Allophones of Standard Canadian English (Boberg 2008:136)

Figure 2.8 illustrates the vowel phonemes and many important allophones of Standard Canadian English within a standard F1×F2 vowel space, transcribed with the same notation discussed above for Table 2.4. The indicated acoustic formant values represent means taken from the acoustic database of the Phonetics of Canadian English (PCE) project (Boberg 2005, 2008), which includes 64 speakers from 6 distinct sub-regions covering the geographic space labelled Canada in Figure 2.6, and 22 speakers from 2 sub-regions comprising Atlantic Canada. As such, this chart should not be taken as precisely indicative of the pronunciations found in any particular region of the country, although it is heavily skewed (74% of speakers) towards the region identified as General Canadian English. Especially with respect to the allophones in Figure 2.8, individual (sub)-dialects will have their own distinctive and particular variants; for phonemes, such as /iy/, for which allophonic variation is not described in the chart, dialectal variation is less significant across regions, and indeed between Canadian and other English dialects.

The primary areas of interest for this dissertation in Figure 2.8 concern the CR diphthongs, transcribed as /ay, aw/ for their non-raised allophones, and /ayT, awT/ when raised. For both diphthongs, the raised allophones are very near to each other, at positions which are both raised and advanced (only slightly in the case of /aw/) from their non-raised counterparts. Boberg measures vowel positions as indicated in Figure 2.8 "at the maximal value of F1 in the case of a vowel whose central tendency is the lowering and raising of the tongue; at a point of inflection in F2 in the case of a vowel whose central tendency is movement of the tongue toward, then away from the front or rear periphery of the vowel space," (p. 134); Boberg does not clarify into which of these two categories the CR diphthongs fall, although presumably it is the former. In any case, the indicated positions represent the maximal value of *one of* the first two vocalic formants, and the value of the other formant at the same position. Interested readers are again directed to Boberg (2010) for further information on other vowels and features of Canadian English.

2.3.3 THE VOWELS OF MANITOBA (WINNIPEG) ENGLISH

This section provides an overview of research on the vowel system of the dialect of English spoken in Manitoba, and specifically the city of Winnipeg, the location of the primary acoustic study carried out for this dissertation. Winnipeg is the capital city of the province of Manitoba, the easternmost of the three Canadian Prairie provinces which also include Alberta and Saskatchewan (Figure 2.3). Aside from Canada's northern Territories, the Prairie region was one of the last parts of the country to be populated by settlers (see §2.3.1) and remains relatively sparsely populated; Manitoba is currently the third-least densely populated province[16] in Canada at 2.2 per sq. km (Statistics Canada 2011), slightly greater than Saskatchewan's 1.8 per sq. km, but well below Alberta's 5.7 per sq. km.

In terms of linguistic research, the Prairies are one of the least studied regions of the continent (again, excluding northern Canada). One of the earliest linguistic publications to examine Prairie English dates back nearly six decades, discussing one researcher's own Saskatchewanian idiolect (Lehn 1959), but linguistic research on the region has generally been lacking. To date there is an almost total absence of in-depth studies even of the major Prairie cities, though see Rosen & Skriver (2015) for an example of more recent work focusing on a specific speech community, Mormons living in southern Alberta.

As of this date there exists, as far as I am aware, a solitary phonetic study based on the vowels of Winnipeg English, Hagiwara (2006) whose stated goal was "to produce quantifiable acoustic baselines for the description of vowels in Winnipeg," (p. 128).

[16] This excludes the sparsely populated federal territories of Yukon, Northwest Territories and Nunavut, which have extremely low population densities of 0.1 per sq. km or less.

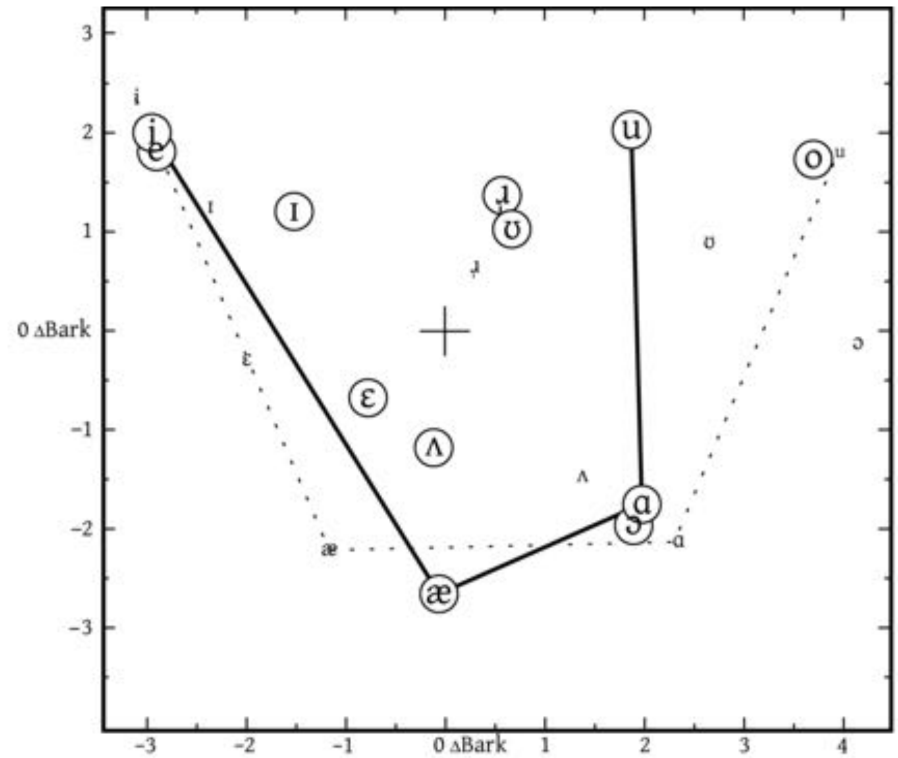


Figure 2.9 Women's vowel centres, Winnipeg vs. California; Bark scale (Hagiwara 2006:132)

Hagiwara's female[17] Winnipeggers' vowels are plotted in Figure 2.9. The Winnipeg vowels are shown in the larger font size, contained within circles, and contrast with vowels from a sample of a demographically similar California (Los Angeles)[18] population (Hagiwara 1995, 1997), shown in smaller font, and uncircled; the heavy solid (Winnipeg), and light dotted (California) lines form open-top trapezoids indicating the "four corners" of the vowel space for each group, respectively.

---

[17] Women's vowels only are provided here for two reasons. First, only women are included in the reported studies in Chapter 4, including Winnipeg, so there are no male counterparts to compare to Hagiwara's male speakers. Second, there appear to be several errors in the reported F2 values for several male vowels in the chart in Hagiwara (2006:131, Table 2) which provided the source data for Figure 2.10, below; when the original values for Hagiwara's men's vowels are plotted, they do not fall in the same location as shown in the related (original) chart.

[18] Hagiwara's participants "had lived all or most of their lives in southern California" (Hagiwara 1995:19).

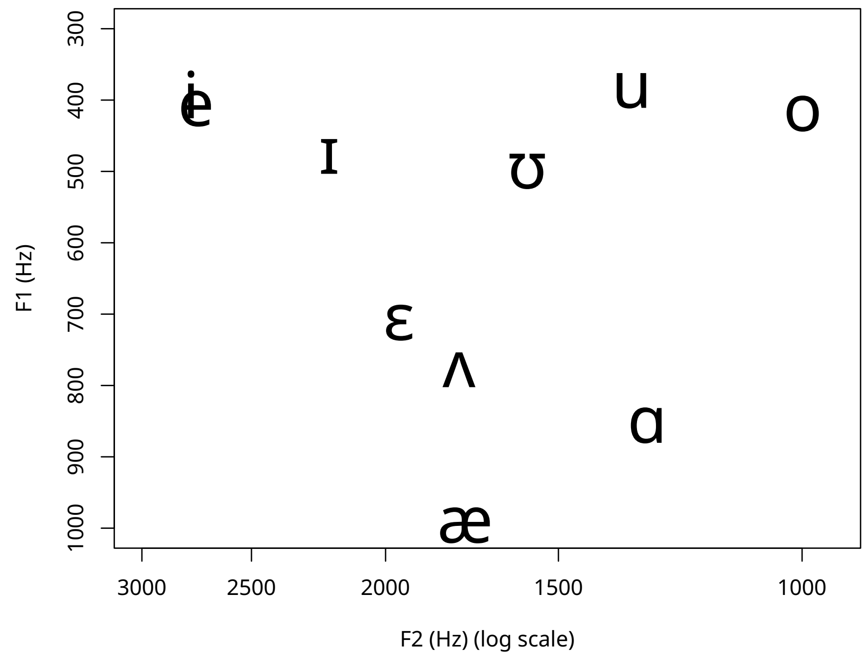


Figure 2.10 Winnipeg women's vowel centres (based on Hagiwara 2006)

The raw data in Hagiwara (2006) was also used to create an alternative plot shown in Figure 2.10. Here, the x-axis F2 values are plotted in a logarithmic scale, which has the effect of expanding the right-edge "back" vowel space relative to the left-edge "front" space (this scaling method is also used in all subsequent vowel plot figures in this dissertation). The resulting plot not only more closely resembles the typical arrangement of a phonological vowel chart, but is remarkably similar to Hagiwara's Bark-scale plot, but without the necessity of being converted from Hertz values, which offers potentially greater comparability with other studies; anecdotally, Hertz values are reported far more frequently than Bark, especially in sociophonetic research (albeit not in logarithmically-transformed values as is done here).

Hagiwara directly compares his Winnipeg results to two American studies, Peterson & Barney's (1952) study of 'General American', and Hagiwara's previous (1995, 1997) study of California speakers. Hagiwara notes the following regarding the Winnipeg vowel system, in comparison to the two American studies:

- **Merger of the low-back vowels, i.e. the 'cot-caught' merger:** in Winnipeg, the merged vowel is phonetically realized as a low-mid vowel; in California, this merger is also present but has a distinct realization which is much lower than the corresponding Winnipeg vowel, and even lower than Californian /æ/, whereas for Winnipeggers /æ/ is the lowest vowel in the system
- **Advancement (centralization) of /u, ʊ, ʌ/:** California (but not General American) has similar realizations as Winnipeg especially for /u, ʊ/, with a somewhat less advanced /ʌ/
- **Retraction and lowering of front lax vowels /æ, ɛ, ɪ/ i.e. *Canadian Shift*** (aka 'California Shift'); Clarke et al. (1995) and others have argued that the low-back vowel merger, also present in both Hagiwara's Winnipeg and California samples, is the impetus for this development, operating as a pull-chain shift[19]
- **Canadian Raising**; Hagiwara provides details on the trajectories of the CR diphthongs by coda context; raised /aj/ is found to follow a generally similar path to non-raised /aj/, with onset and offset occurring higher and further forward for the former; /aw/ is somewhat different, with the entire trajectory substantially raised but not especially retracted; Hagiwara does not contrast the Winnipeg diphthongs

[19] Hagiwara notes certain dissimilarities between his and Clarke et al.'s findings and concludes that "Canadian Shift, as described by Clarke et al. (1995), does not seem to characterize either the Winnipeg or the Southern California samples," (Hagiwara 2006:136); however, as Canadian Shift is not a focus in this dissertation, I will not explore the topic further.

with his American data, so any differences between the two are unknown; see §4.3 for comparison of Hagiwara's data with results from the present study

Although Hagiwara (2006) represents the only previous work documenting the properties of the entirety of the vowel system of Winnipeg English, Onosson (2010) conducted an examination of the CR diphthong /aj/ which focused on durational patterns in raising and non-raising contexts. In that work, duration by coda context was measured, which provides a PVVA ratio of 0.539 (see Table 4.5, below). Statistical testing of acoustic qualities including formant values and overall duration indicated that the largest significant difference between raised and non-raised /aj/ concerned the factor of duration rather than F1, which is associated with vowel height, and that F2, associated with vowel advancement/retraction, was also a significant factor related to the observed variance between the two sets.

2.4 ENGLISH IN THE (AMERICAN) WEST

In this section, the history and contemporary characteristics of the English dialect known as The West, spoken in the area around Denver, Colorado, are described. The primary sources consulted here are Ubbelohde, Benson & Smith (2006) and Abbott, Leonard & Noel (2013) for the general history of Colorado (§2.4.1), and Clopper, Pisoni & de Jong (2005) and Labov et al. (2006) for the description of contemporary English in Colorado (§2.4.2).

2.4.1 A BRIEF LINGUISTIC HISTORY OF COLORADO

The pre-Anglophone history of Colorado is complex; multiple groups existed in and laid claim to the region at various times. Prior to European exploration of the area, groups identified by the terms Ancestral Puebloan (Anasazi) and Frémont occupied Colorado. Subsequent peoples present in the area at the time of European contact include the Ute, Apache, and Navajo, joined later from neighbouring regions by the Cheyenne, Comanche and Arapaho among others.

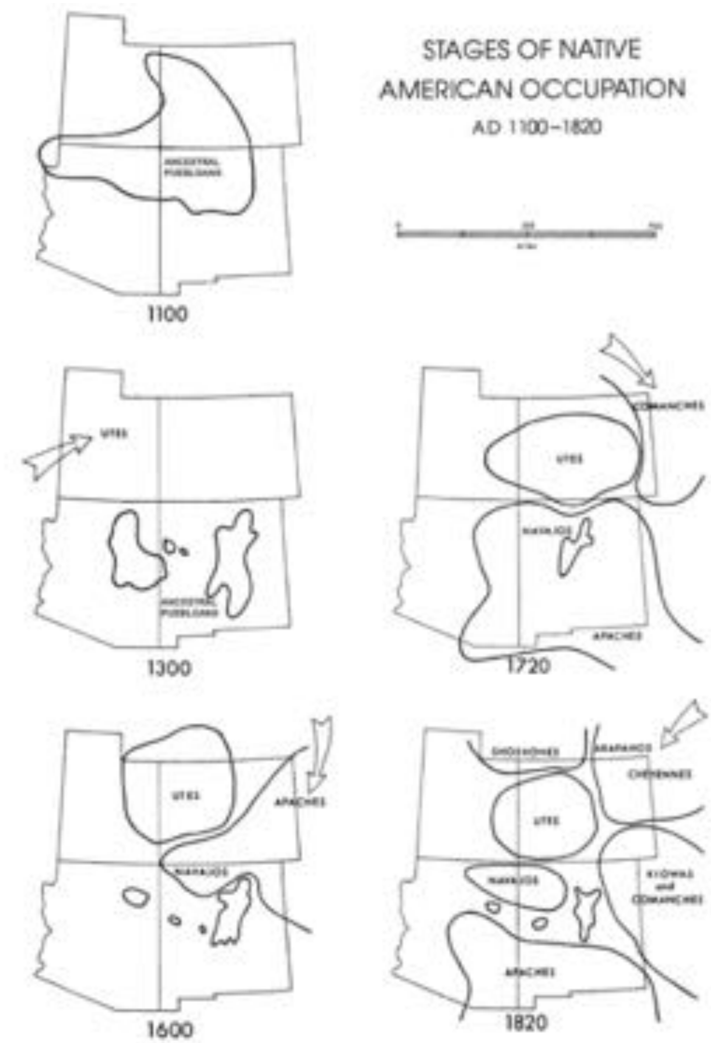


Figure 2.11 Stages of Native American Occupation (Abbott et al. 2013)

Figure 2.11 displays the areas occupied by the major Indigenous groups of Colorado at various points in time up until 1820; Colorado is the upper-rightmost among the four states (which also include Utah, Arizona, and New Mexico) pictured in the maps. At the end of the 16th century and continuing throughout the 17th century, Spain initiated settlement within its territory of New Mexico, immediately to the south of Colorado. By the time the Spanish began making a concerted effort to settle Colorado itself in the late 17th century, they found French traders already present in the region having arrived from New France to the east, although no permanent French settlements had been established. Skirmishes between proxy groups occurred as the French and Spanish battled for supremacy, with France eventually ceding claim in 1763, only to retake the area in 1800 before shortly thereafter retreating almost entirely from the North America, turning over most of their territory (including parts of Colorado) to the United States as part of the Louisiana Purchase in 1803.

The Louisiana Purchase gave the United States possession of a portion of the region comprising modern Colorado, although the borders with New Mexico were unclear at the time; a treaty in 1819 settled the border disagreement and placed most of present-day Colorado under Spanish control. American exploration of the area began in late 1806, but was sporadic throughout the early 19th century until the conclusion of the Mexican-American war in 1848. The war ended with the entirety of Colorado as well as all of the surrounding regions falling firmly under American control. After the conclusion of the war, a sizeable number of Spanish-speaking colonists remained in Colorado but found their language, culture and religion increasingly under threat from Anglophone settlers.

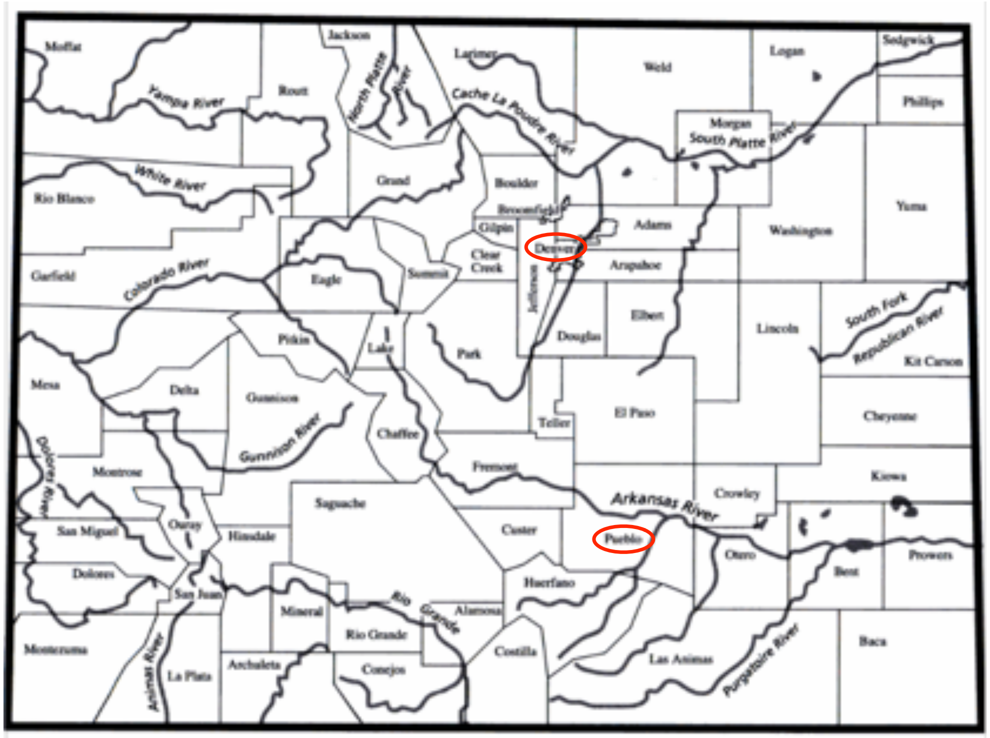


Figure 2.12 Colorado's Major Rivers and Counties (Ubbelohde et al. 2006:xvii)

An early American trading fort had been established in 1833 near modern-day Pueblo, less than 200 km south of Denver (see Figure 2.12; the locations of Pueblo and Denver are highlighted in red). American settlement in the area greatly expanded beyond Pueblo after the conclusion of the war, spurred on by a gold rush in the late 1850s which brought an influx of English-speaking Americans from the eastern states. Spanish as a spoken language in Colorado receded from this point on, giving way to English, although descendants of the early Spanish settlers still reside in Colorado to the present day (several of the participants recruited in Denver for this study identified themselves as having such ancestry). Colorado attained statehood in 1876, effectively formalizing its relationship with the anglophone United States.

At present, Colorado is overwhelmingly anglophone, with 83% of the population being monolingual in English. A sizeable (home-language) Spanish-speaking community of approximately 584,000 people exists (U.S. Census Bureau 2015), but this is composed of recent migrants to the area rather than the earlier Spanish settlers. No Indigenous (pre-European-contact) languages are spoken widely in Colorado at present.

2.4.2 The vowels of The West

The information presented in this subsection focuses exclusively on the vowel system of The West, the dialect spoken in Colorado, with special attention paid to the diphthongs and other known areas of contrast with Canadian English (see §2.3). While most of the findings are drawn from Labov et al. (2006), additional comparative data are presented from Clopper, Pisoni & de Jong (2005), which was published while the former work was being completed, and which is one of the first attempts to "provide a summary and description of the acoustic characteristics of the vowel systems of [multiple] regional varieties of American English," (p. 3) including The West and The North.

In their comprehensive account of North America varieties of English, Labov et al. (2006) present a goal of defining and demarcating English dialects in a principled manner derived from acoustic and phonological analysis and comparison (p. 119). These same principles which were used to identify and define the region of Canadian English (§2.3) also identify and define the region which includes the city of Denver, known simply as *The West*.

Figure 2.13 The West and its neighbors (Labov et al. 2006:280)

The West is the largest regional dialect of North American English from a geographical point of view, covering major portions of at least eleven, and smaller areas of another three to five, U.S. states; essentially, it comprises the western one-third of the continental U.S. as shown in Figure 2.13. The West covers such a vast area due to the fact that, like much of the region where Canadian English is spoken, it was settled relatively late compared with other English-speaking parts of North America, and fairly rapidly (see §2.4.1). Unlike Canada's east-to-west scope, the geographic expansiveness of The West extends in all cardinal directions, which speaks to its internal diversity. Labov et al. (2006) note that The West "lacks the high levels of homogeneity and consistency that was found for most other dialects" and that while its unique identification is possible, as "The West shows trends or tendencies that differentiate it from its neighbors … many of its characteristic features are also found in quite distant regions," (p. 284).

The most important identifying features as described by Labov et al. (2006) for The West are described as follows (phonological transcriptions follow the format used by Labov et al. 2006, as described in Table 2.4):

- **Fronting of /ow/:** The West in general exhibits fronting of /uw/ but not /ow/; however, this is not homogenous throughout the region. Denver in particular is indicated as having moderate fronting, with F2 of /ow/ between 1200-1400 Hz, compared with non-fronting regions (e.g. Canada) where it is below 1100 Hz, and more extreme fronting regions (e.g. The South) where it exceeds 1400 Hz.
- **Fronting of /aw/:** The position of the nucleus of /aw/ is also a variable feature in The West, although it is generally more fronted than in Canada or the North, with F2 greater than 1450 Hz throughout the region, and frequently exceeding 1650 Hz. However, comparing the cities of Winnipeg and Denver specifically, Figure 2.14 indicates a higher level of similarity, with at least some Denver speakers falling within the same range as Winnipeggers, who have more fronted /aw/ nuclei than some other Canadians, especially in other parts of The Prairies; note, too, that at least part of the sampled population in Madison also falls within the F2 > 1450 Hz range for /aw/. Note: it is unspecified whether or not the data for this analysis included both/either voiced and voiceless codas, although given analysis of /aw/-raising in other parts of Labov et al. (2006) it seems more than likely that both were involved.

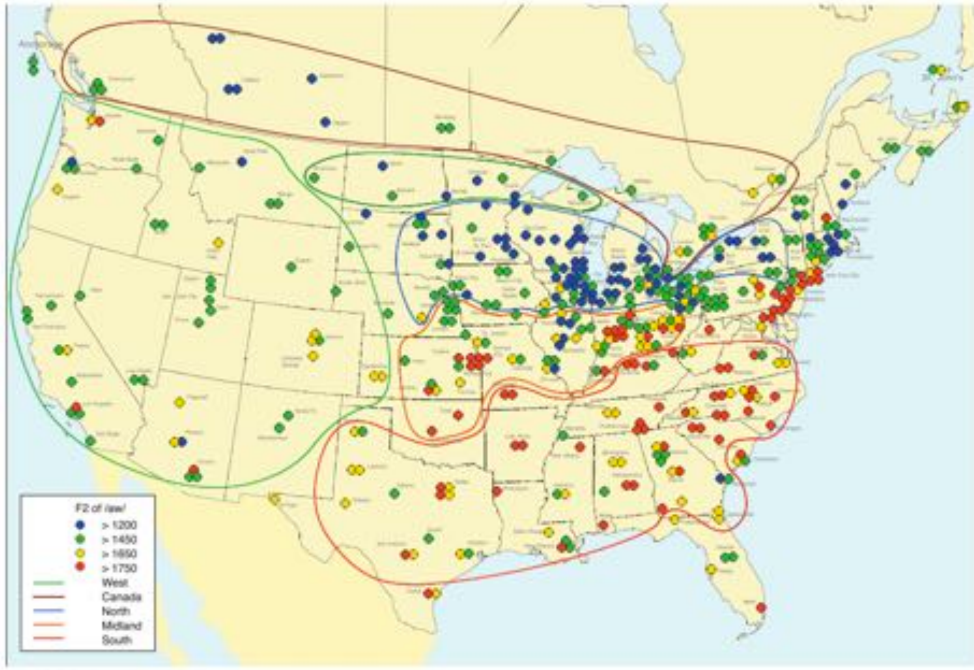


Figure 2.14 Fronting of /aw/ in the West (Labov et al. 2006:282)

- **Raising, fronting of front tense vowels:** The mean positions of the front tense vowels /iy/ and /ey/ in The West are among the highest and frontest in the U.S., but somewhat less so in both respects than Canada.
- **Absence of Canadian Shift:** As stated, The West is characterized by a general lack of strong participation in the Canadian Shift, defined broadly as retraction and lowering of front lax vowels /æ, ɛ, ɪ/ (Clarke et al. 1995); this is not the case with the similar California Shift (Eckert 2008; Aiello 2010; Kennedy & Grama 2012)[20] which occurs within the sphere of The West, but which is geographically far from the sampled population in Denver.

[20] As discussed previously, given that Canadian/California Shift does not form one of the foci of this dissertation, I express no particular view on whether these terms represent one process (with regional differences) or distinct phenomena.

- **Absence of Canadian Raising:** Defined as the co-occurrence of pre-voiceless raising of the nuclei of both /ay/ and /aw/, Labov et al. (2006) find CR to be largely absent within The West.

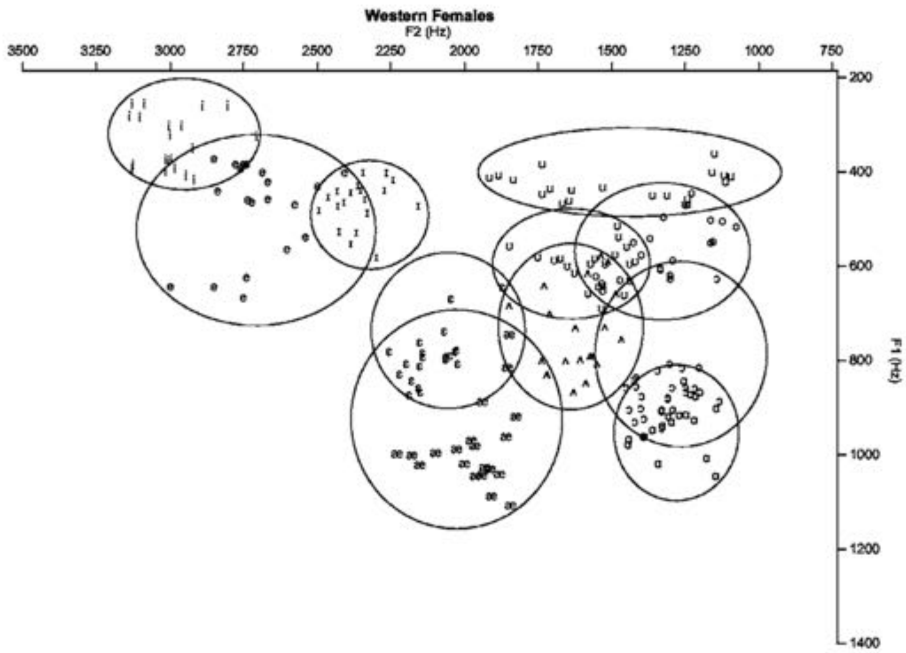


Figure 2.15 Vowel tokens of Western Females (Clopper et al. 2005:27)

Figure 2.15 presents the vowel tokens of four female speakers from The West in Clopper et al. (2005); as elicitations in that study were exclusively based on the frame *hVd* e.g. *head, had,* etc., these tokens indicate the acoustic characteristics of vowels occurring before voiced codas only. Some of the features described by Labov et al. (above) and visible in Figure 2.15 include fronting of /i, e, o/. However, as Clopper et al. did not include diphthongs in their samples, no comparisons can be made with regard to /aw/-fronting or CR occurrence.

2.5 English in The (American) North

In this section, the history and contemporary characteristics of the English dialect known as The North, spoken in the area including Madison, Wisconsin, are described. The primary sources consulted for this section are Smith (1973), Nesbit (1973, 1976) and Current (1977) for the general history of Wisconsin (§2.5.1), Purnell, Raimy & Salmons (2013) on its specifically linguistic history (also §2.5.1), and Labov et al. (2006) for the description of contemporary Wisconsin English (§2.5.2).

2.5.1 A brief linguistic history of Wisconsin

Prior to European arrival, the region including present-day Wisconsin was inhabited by several archaic cultural groups, with uncertain connections to later cultures. Despite its small size, by the time of European contact and the first written histories of the area, the population of Wisconsin and the surrounding region included a wide variety of different peoples and cultures, some of the most prominent or well-known among these being the Ho-Chunk (Winnebago), Huron, Chippewa (Ojibwe), Sauk (Sac), Fox, Miami, and Menominee. In the early 1600s, Samuel de Champlain, the founder and de facto governor of New France, enacted a policy of sending individuals into the interior of North America to learn the language and culture of the groups with which the French were engaging in trade. The earliest European individual to enter what is now Wisconsin was most likely one of a handful of French fur-traders sent by Champlain in the 1620s under this practice.

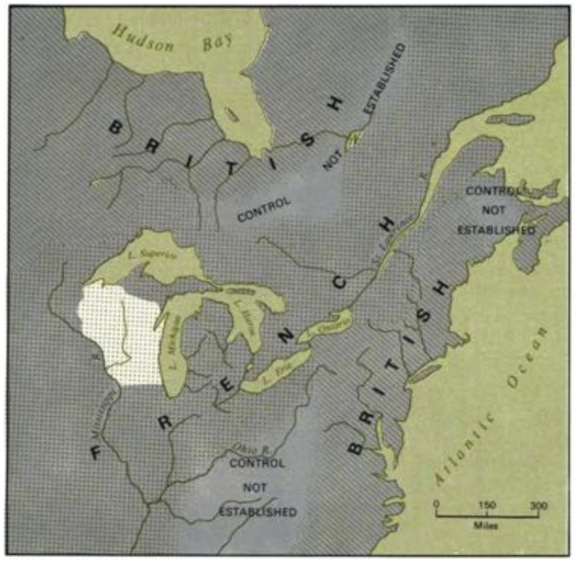
Trade in beaver pelts, which brought the French to the region, was critical in the history of Wisconsin. As beaver populations became depleted in the lands around the eastern Great Lakes, groups of Iroquois peoples moved west in the 1640s into the area including Wisconsin and drove out resident Huron populations and their allies in order to claim their more prosperous territory. The following decades were quite turbulent as a variety of groups entered or left the state, or relocated internally in the wake of these attacks and their aftermath.

Figure 2.16 Indian Tribes of the Western Great Lakes (Smith 1985:15)

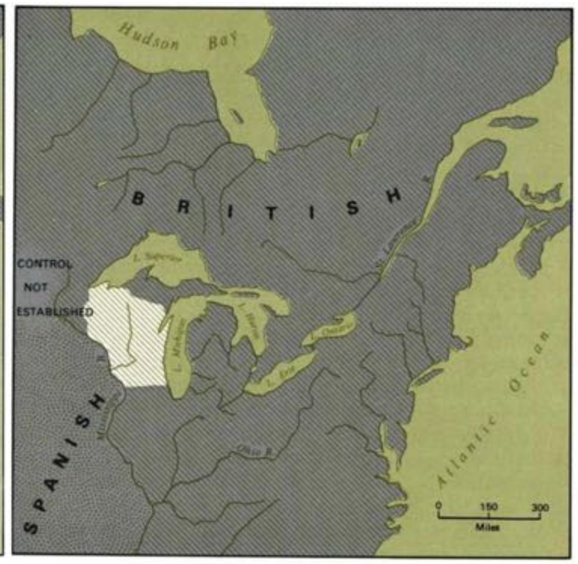
Figure 2.16 illustrates the locations of a number of Indigenous groups in the western Great Lakes region throughout the eighteenth and nineteenth centuries; present-day Wisconsin is centred within the area of the map, occupying most of the land from the coast of Lake Michigan westwards, and to the south of Lake Superior, excluding the Lake Superior coastline except the southwestern-most portion.

Among European peoples, the French were the sole or dominant presence in the Great Lakes region throughout the seventeenth century, building dozens of trading forts throughout the area during that time, although not settling in the area in large numbers. By the 1680s the French had established a vast trading network that extended from Québec through the Great Lakes and down the Mississippi to the Gulf of Mexico, and claimed sovereignty over the entire region, with Wisconsin forming a central hub for the network.

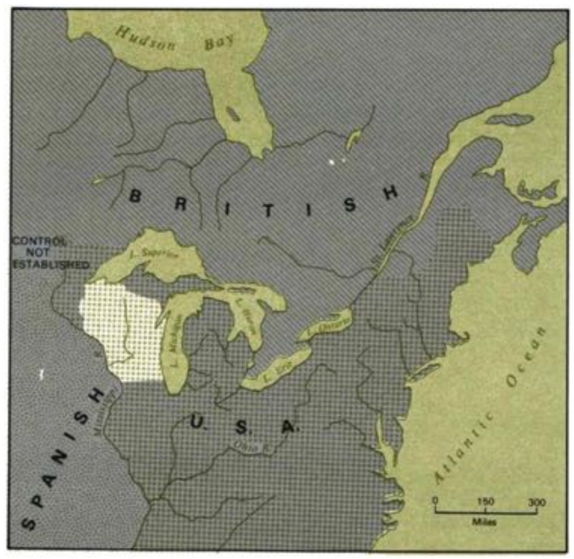
France's domination of the North American interior was not to last, however. The British began encroaching on the region from both the south via their American colonies, and from the north following the establishment of the Hudson's Bay Company in 1670. A long series of conflicts involving France, Britain, Spain, and their respective Indigenous allies throughout the eighteenth century led to France almost completely ceding its North American territory, with Wisconsin eventually falling under British control under the Treaty of Paris in 1763, and subsequently to the United States following the American Revolution and the establishment of the border between British North America (Canada) and the United States, which mainly followed the Great Lakes and the St. Lawrence River. The various changes in territorial possession of eastern North America by the colonial powers throughout the eighteenth century is illustrated in Figure 2.17, with the location of present-day Wisconsin highlighted.



SPHERES OF INTEREST, 1713



SPHERES OF INTEREST, 1763



SPHERES OF INTEREST, 1783

Figure 2.17 Spheres of Interest: 1713, 1763 and 1783 (Smith 1985, pp. 41, 54, 72)

It was thus only at the end of the 18th century that an Anglophone presence began to exist in Wisconsin. The fur trade continued to dominate the economic activity of the region and the British (Canadians), through the North West Company, continued to be the major traders in the American Great Lakes region even into the nineteenth century despite the change in sovereignty. The War of 1812 between Britain and the United States ended this situation as far as Wisconsin was concerned, following which most British subjects subsequently left for or were expelled to Canada. Conflict with Indigenous groups inevitably followed, as the United States sought to solidify the (European-origin) American presence in the area. While some groups such as the Menominee and the Oneida remain in Wisconsin to this day, others such as the Ho-Chunk and Sauk were forcibly removed or left the area following battles with American forces.

From roughly the turn of the 19th century, European immigration to Wisconsin brought a range of languages to the state, frequently including Germanic languages such as German, Norwegian and Dutch (Lucht 2013:26). Immigrants constituted a large portion (up to one-third) of the population of Wisconsin mid-century, mainly Germanic or anglophone until towards the latter decades when southern and eastern European immigrants began to arrive in more substantial numbers. The earlier pattern of predominantly Germanic immigration, including formal non-anglophone education, has been argued to have some lasting effects on the phonology of Wisconsin English into the present-day, especially concerning the expression of post-vocalic voicing distinctions, which are not maintained in the German language (Purnell, Salmons & Tepeli 2005; Purnell, Salmons, Tepeli & Mercer 2005; Petty 2013). This is an important factor regarding the present study, as such distinctions are relevant to the expression of PVVA effects (§2.1). Of further relevance, Madison, one of the cities selected for recruitment of participants (Chapter 3) forms one of the vertices of a triangular region identified as an area of high German-language influence on present-day Wisconsin English.

At present, Wisconsin is overwhelmingly Anglophone (and more so than both Manitoba and Colorado), with more than 91% of the population being monolingual in

English. A sizeable (home-language) Spanish-speaking community exists, numbering 247,000 (U.S. Census Bureau 2015), composed of fairly recent migrants. At present, no Indigenous languages are spoken in large numbers in Wisconsin.

2.5.2 THE VOWELS OF THE NORTH

This section presents a contemporary view of the vowel system of The North, the dialect spoken in southern Wisconsin, with emphasis placed on the diphthongs and other areas of contrast with Canadian English (see §2.3). Unlike The West (see §2.4.2), Labov et al. (2006) identify a more homogeneous space for The North, albeit one with some internal divisions.

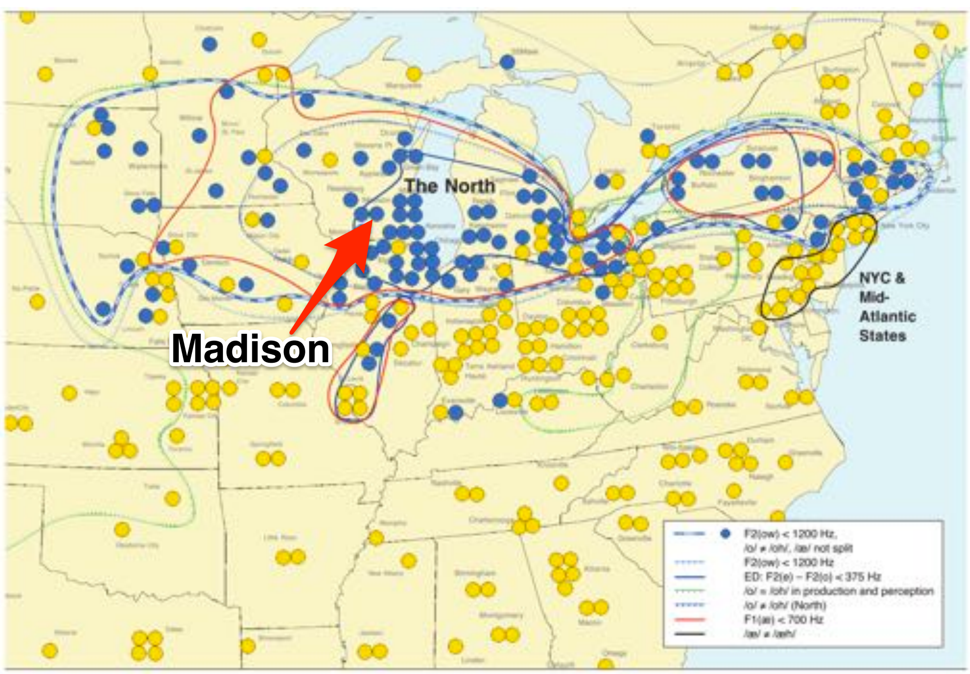


Figure 2.18 The outer limits of the North (Labov et al. 2006:134)

Figure 2.18 illustrates the main isogloss boundary defining The North, as well as some of these internal divisions as indicated in the map legend[21]. The position of Madison is indicated with an overlay; the city falls within the isogloss boundaries of all the core Northern features, a subregion known as the Inland North (see also Figure 2.6), albeit very near its western edge. The outer boundaries of The North, distinguishing it from neighbouring dialects, can be described by a variety of important features; I focus here on those features having notable overlap with, or distinction from the characteristics of Canada and The West:

- **Lack of merger of /o/ and /oh/:** This refers to what is also known as the *cot-caught* merger; in this respect, The North is distinct from both Canada and The West, in which regions the merger is largely or totally complete.
- **No /ow/-fronting:** F2 of /ow/ typically falls below 1200 Hz. This is a property that The North shares with most of Canada (excluding British Columbia), but differentiates it from The West which exhibits higher values, exceeding 1400 Hz in many areas, although this feature is quite variable throughout The West.
- **Back position of /aw/ relative to /ay/:** Figure 2.19 illustrates an area bound by the *AWY line*, within which F2 of /aw/ is lower than F2 of /aj/, i.e. the nucleus of /aw/ is further back than that of /aj/. The AWY line not only surrounds the majority of The North, but extends into the Prairie region, including Winnipeg (but excluding other regions of Canada). In most other dialects, including The West (and Denver in particular), /aw/ is further forward (higher F2) than /aj/.

[21] For those features not elaborated upon below, see Labov et al. (2006), Chapters 11 and 14.

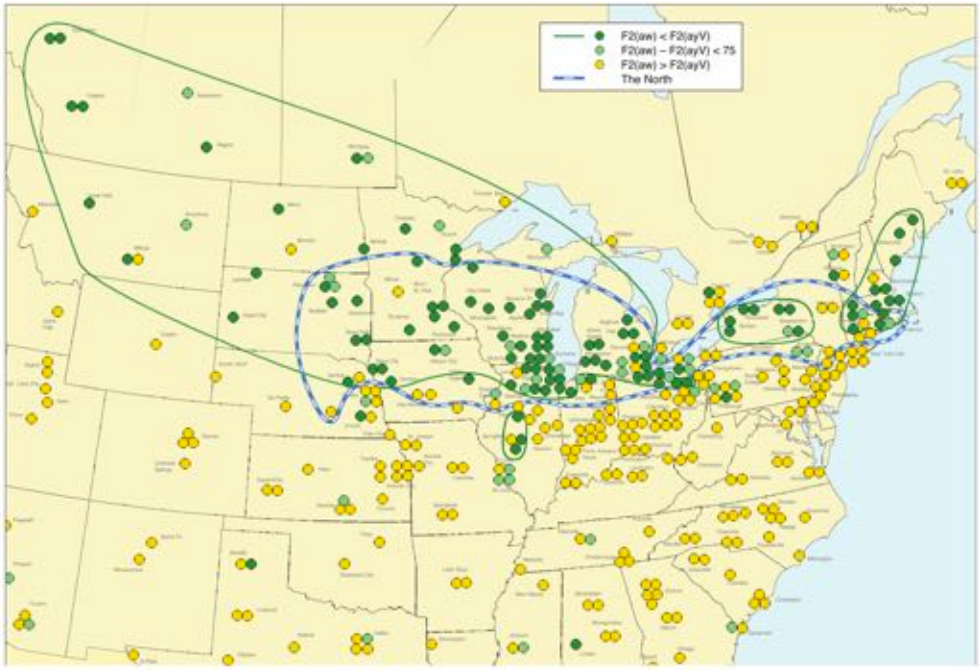


Figure 2.19 Relative fronting of /aw/ and /ay/ and the AWY line (Labov et al. 2006:188)

- **Raising and fronting of /æ/, i.e. *Northern Cities Shift* (NCS):** The NCS is a complex pull-chain shift[22] involving seven vowels taking place in the Inland North subregion, thought to have been initiated by raising and fronting of /æ/, as illustrated in Figure 2.20; the city of Madison is shown in Labov et al. (2006) to exhibit all of the NCS features. Both /æ/-raising and the NCS are indicated via isogloss boundaries in Figure 2.18 above: the /æ/-raising region is bound by a solid reddish orange isogloss, denoted as *F1(æ) < 700 Hz* in the legend; the NCS region is bound by a solid bluish-purple isogloss, *ED: F2(e) – F2(o) < 375 Hz* in the legend. Both features are absent in Canada and The West.

---

[22] There may be other interpretations of NCS which take it to be several distinct processes instead; I stake no claim here in describing it as one entity, and the term may be thought of as simply a reference to the co-occurrence of this particular cluster of features which, following Labov et al., can be ascribed to the region including Madison.

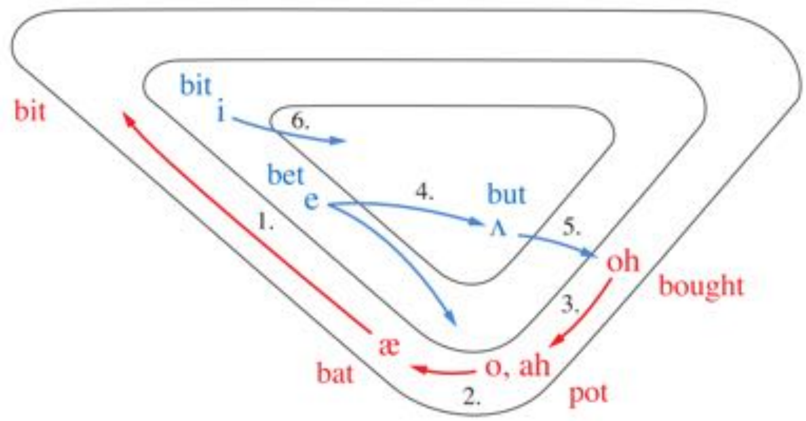


Figure 2.20 The Northern Cities Shift (Labov et al. 2006:121)

- **Raising of /ay/:** Labov et al. (2006) define /ay/-raising acoustically as a 60 Hz decrease in F1 before voiceless consonants vs. other environments. It is found fairly commonly but quite variably throughout The North, although, as illustrated in Figure 2.21, it occurs (as defined by this criterion) variably to some extent in nearly every region of the U.S. Generally speaking, however, regular occurrence of /ay/-raising within the U.S. is concentrated in The North and the neighbouring North Central regions; the latter, spanning from eastern Montana through to the western part of Michigan's Upper Peninsula and mostly adjacent to The Prairies, is the only U.S. region where /ay/-raising is ubiquitous.

Figure 2.21 Canadian raising of /ay/ (Labov et al. 2006:206)

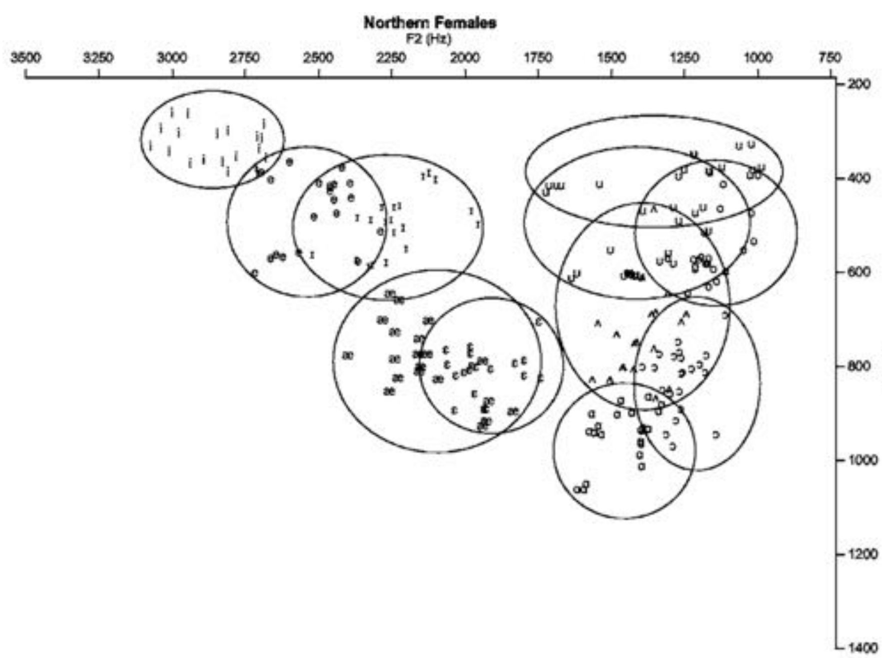


Figure 2.22 Vowel tokens of Northern Females (Clopper et al. 2005:24)

Figure 2.22 illustrates the vowel tokens of four female Northern speakers investigated by Clopper et al. (2005); as with Figure 2.15, all tokens were elicited from words in the frame *hVd*, and so only represent the characteristics of vowels before voiced codas. Additionally, as with Figure 2.15, no diphthongs were analyzed by Clopper et al. Many of the features described by Labov et al. (excluding the diphthongs) for The North are evident in Figure 2.22 including: lack of /ɑ~ɔ/ merger, lack of /o/-fronting, and raising and fronting of /ae/, with many tokens exceeding /ɛ/ in both respects.

2.6 Cross-dialect summary: Canada, The West & The North

Here a brief summary is presented of the findings in sections §2.3.3, §2.4.2 and §2.5.2. In Table 2.5, a number of features which were highlighted within one or more of the examined dialects are listed, and their presence or absence within each dialect, on the basis of literature discussed in the preceding sections, as well as comparison of vowel charts.

Table 2.5 Cross-dialect comparison of phonetic features

Feature	Canada	The West	The North
/aj/-raising	✔	✗	✔
/aw/-raising	✔	✗	✗
/aw/-fronting	✗	✔	✗
/u/-fronting	✔	✔	✗
/o/-fronting	✗	✔	✗
Merger of /ɑ, ɔ/	✔	✔	✗
Canadian Shift; lowering and retraction of /æ, ɛ, ɪ/	✔	✗	✗
Northern Cities Shift; raising, advancement of /æ/	✗	✗	✔

None of the features presented in Table 2.5 has unanimous patterning across all three dialects, neither is there a uniform split of one dialect vs. the other two. In some cases, Canada and The West stand alike vs. The North, e.g. the /ɑ, ɔ/ merger and /u/-fronting; for other features it is Canada and The North vs. The West, e.g. presence of /aj/-raising or lack of /aw/-fronting; and in still other cases, The West and The North fall into line together vs. Canada, e.g. /aw/-raising and Canadian Shift. Each dialect included here thus represents a unique bundle of features. Due to this three-way patterning across diverse features, as well as their geographic proximity to each other, this particular set of dialects offers a suitable venue for examination of any of the indicated features; within this dissertation, the focus is on the features relevant to CR, namely /aj/- and /aw/-raising. In the following chapters, the collection and analysis of data from each indicated city is examined, focusing closely on differences in CR patterning between the three dialects.

Chapter 3 Methodology

This chapter describes the methods used for three connected studies[23] which were carried out between 2014–2017 in the three cities: Winnipeg, Manitoba, Canada; Denver, Colorado, U.S.A.; and, Madison, Wisconsin, U.S.A. The methods of data collection and analysis were virtually identical across all three studies, and are described generically here, along with relevant additional comments for the individual studies where necessary.

The first study was conducted on a population sample from Winnipeg, Manitoba, Canada between April of 2014 and June of 2015, and is intended to be representative of the geographically broad variety of Canadian English spoken across Central and Western Canada. The second and third studies were carried out in cities chosen to be complementary to the Canadian dialect with respect to the articulation of the diphthongs /aj, aw/; that is, with respect to the production of Canadian Raising (CR). The second study was carried out in Denver, representing the dialect of The West where CR is described as not generally occurring (see §2.4.2). The third study was carried out in Madison, representing the dialect of The North where raising of /aj/, but not /aw/, is known to occur (see §2.5.2). Data collection for both the Denver and Madison studies was carried out in May of 2017.

3.1 DATA COLLECTION

The participants for each study were recruited in their respective cities or regions. All participants are native speakers of English, residing in the city (or neighbouring

[23] All data collection for this dissertation was carried out between 2014–2017 under approval from the Human Research Ethics Board of the University of Victoria, protocol #14-106.

communities) pertaining to each study, and born between 1954 and 1996. Both male and female subjects were recruited in Winnipeg, as the study was originally planned to concentrate solely on that speech community. When the scope of the study expanded and it became desirable to add other communities for comparative purposes, only female subjects were recruited beyond Winnipeg, and the male Winnipeg data was set aside for future examination. As recruitment outside Winnipeg took place partially while in the field, there was limited time within which to recruit participants, travel to meet them (all recordings were conducted at participants' homes or workplaces) and conduct recordings.

In order to maximize the efficiency of this process, the choice was made to restrict the populations being compared to one sex. This allowed a greater number of participants of similar demographic makeup to be recruited, rather than having more demographic types with fewer numbers of each type, had males also been recruited. There are also two advantages forthcoming from restricting the analysis to only female voices. First, non-normalization of formant values becomes less problematic, due to the greater degree of similarity in the expected length of same-sex vocal tracts. Second, as female voices tend to produce formant values which cover a wider span (in Hertz), dynamic movement through the vowel space, such as produced by diphthongs, may be easier to detect.

Table 3.1 Breakdown of study participants' ages

Age (at interview date)	Winnipeg	Denver	Madison
50–59	2	4	4
40–49	10	4	0
30–39	4	4	2
20–29	4	3	3
Total, N =	20	15	9
Mean age	37.9	39.3	42.1

Table 3.1 presents the demographic makeup of each study cohort. As can be seen, recruitment in Madison was especially problematic, and there was difficulty in obtaining ideal quantities of speakers of certain ages. Conversely, in Winnipeg there was an abundance of participants ages 40–49; this is unsurprising, as the author falls in this age bracket, and recruitment in Winnipeg took place partly through social networks via word of mouth. Ideally, the three populations would be more balanced across cities and ages, but it is hoped that this lack of balance will not unduly affect the conclusions drawn from comparing the sampled populations.

Arrangements were made with each participant to meet them in a convenient location to conduct recordings. In most instances, the participant selected their residence, while a few chose to use an available private space at their place of employment. In all cases, attention was paid to ensuring that background noise was minimal during the recording process, and it never compromised analysis of the resulting audio files. Each session took no more than thirty minutes, from start to finish. Audio recordings were made directly in Praat (Boersma & Weenink 2016)[24] running on an Apple MacBook Pro computer (model A1278), using an Apogee One combination condenser microphone and analog-to-digital converter, and saved as 44.1 kHz, 16-bit .wav files. The microphone/converter was mounted on a small tabletop microphone stand, placed approximately 12 inches in front of the subject at roughly the same height as their mouth. Words from a prepared list (see Appendix A) were displayed on the computer screen, one at a time, in plain white text on

---

[24] Recordings made in 2014 (Winnipeg) used Praat version 5.3, those made in 2015 (Winnipeg) used version 5.4, and those made in 2017 (Denver and Madison), as well as audio analysis of all the audio files was conducted using version 6.0. I am aware of no reason to expect that the use of different software versions should have affected any of the results.

a solid black background, for high contrast and visibility and to reduce any distractions; nothing else was visible on the screen. The order of presented words was fixed across speakers. This order was originally randomly-generated, and then modified by separating same-vowel words so they did not appear adjacent to each other. I felt that this particular method of ordering was ideal due to the large quantity of words representing just the three diphthongs /aj, aw, ɔj/, as truly random ordering would result in successive words containing the same vowel with high frequency. Each word appeared for approximately three seconds before the next word was displayed. Participants were asked to read each word aloud twice, to mitigate technical or pronunciation errors. Both tokens (barring errors; see below) were retained for all speakers in subsequent analysis.

The wordlist consisted almost solely of monosyllables, and three bisyllabic words with primary stress on the target syllable: *adroit, avoid* and *oblige*. The primary aim of the wordlist's design was to elicit the three English diphthongs /aj, aw, ɔj/ in as many contexts as possible, with the remainder of the vowel inventory being elicited in a smaller set of contexts. The diphthong portion of the list was based on Hammond's (1999) overview of the legitimate phonotactic combinations in which the diphthongs occur; all legitimate diphthongal monosyllables identified by Hammond were included in the wordlist. In addition, all legitimate (as well as some nonce) combinations of English vowels in the frames /h__#/, /h__d/, and /h__t/ were also included to elicit the full range of English vowels in three specific contexts: no coda (i.e. open syllable)[25], voiced coda[26], and

---

[25] Although this context is prohibited in English for most lax vowels, e.g. */hɪ/ */hɛ/, the words *hah* /hæ/(?) and *huh* /hʌ/ were included as an attempt to elicit /æ, ʌ/ in open syllables.

[26] During the construction of the elicitation script for the Winnipeg audio recordings, the word *hood* was accidentally omitted, leading to an absence of tokens of the vowel /ʊ/ before a voiced coda. This vowel was

voiceless coda, with the initial /h/ serving to minimize coarticulatory effects between the onset consonant and the target vowel. In some cases, subjects did not produce the desired pronunciation[27], e.g. pronouncing *lithe* as [lɪθ] instead of [lajθ] or [lajð] (both of which were retained when they occurred); such non-target pronunciations were excluded from further analysis. The full set of 84 items included in the wordlist are provided in the Appendix.

3.2 SEGMENTATION

Segmentation of speech sounds is an essential component of acoustic analysis within phonetic research. As House & Fairbanks (1953) note, "[t]he identification of the beginning and end of a vowel surrounded by consonants is an arbitrary act that is both difficult and artificial," (p. 107). With that caveat in mind, this section details the procedures used in carrying out this task, for purposes of clarity and reproducibility.

Demarcation of vowel boundaries was conducted in Praat. This was done strictly manually in order to ensure maximum fidelity and accuracy. Regions of the visible audio waveform and/or spectrogram representing vowels were identified on the basis of several visual factors including, in relative but not strict order of importance:

- The occurrence of clear and steady spectral formants as visible in the spectrogram, and contrasting with adjacent regions

---

therefore omitted entirely from all subsequent analysis for Winnipeg. This omission was corrected prior to data collection in Denver and Madison, so those datasets contain enough data for complete analysis of /ʊ/; however, for purposes of comparison with the Winnipeg data, it is excluded in most of the analysis reported below.

[27] The vast majority of such 'mispronunciations' can be attributed to a handful of words, largely nonce forms, listed here with their desired pronunciation(s): *bowed* [bawd] (not [boʊd]); *coif* [kɔjf] (not [kwɑf]); *haw* [hɑ]; *hote* [hoʊt]; *houst* [hawst]; *hoyed* [hɔjd]; *hoyt* [hɔjt]; *hud* [hʌd]; *lithe* [lajθ,lajð] (not [lɪθ]); *mouthe* [mawð].

- Glottal periodicity, as evidenced by regular pulses in the waveform
- Visible pitch contour and the presence of a “voice bar” in the spectrogram
- Relatively high intensity compared to adjacent regions, visible in both the waveform and spectrogram

For syllables with an onset /h/ and a coda /t, d/ or an open syllable, demarcation following this protocol was fairly straightforward, due to the relative distinctiveness of the adjacent segments, and the fact that /h/ was preceded by a pause in the elicitation sessions (as participants were instructed to do). Vowel regions were highlighted in Praat, with the locations of the onset and offset marked at the closest zero-crossing within the waveform. Implementation of the segmentation protocol described above is exemplified in Figure 3.1. Visual analysis was always accompanied and augmented by audition of the waveform to confirm the appropriateness and validity of the segmental boundaries.

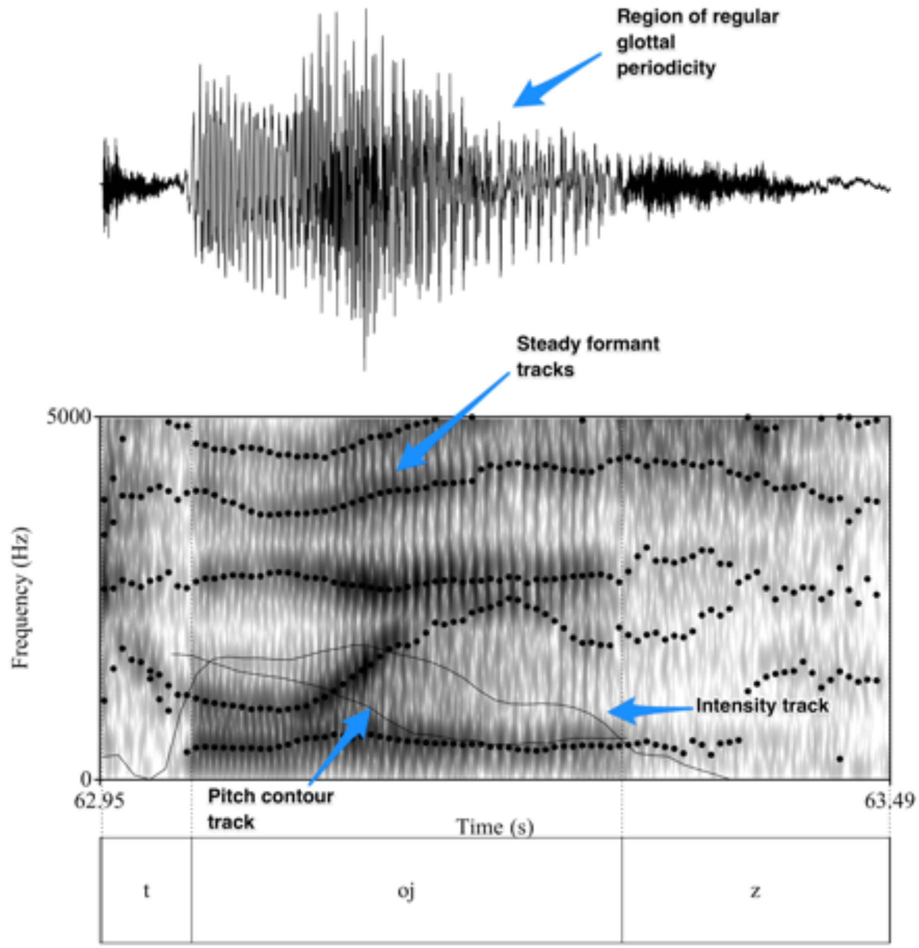


Figure 3.1 Standard vowel tagging protocol: 'toys', speaker AK69f

A few highly sonorant onset and coda consonants—the liquids /ɹ, l/ and the nasals /m, n/—required slightly different individuated protocols, described as follows. Note that in all of these cases, due to the structure of the elicitation wordlist, the only vowels involved were the diphthongs: /aj, aw, ɔj/. The general guiding principle behind each of these protocols was to demarcate the vowel (diphthong) according to the most extreme values found for F1 (related to articulatory height) when considering the nucleus, or F2 (related to articulatory front-to-back position) for the off-glide, relegating the majority of the transitional motion within these formants to the consonant, where this differed from the general path of the diphthong during its central phase. For example, for a typically downward-sloping trajectory such as F1 of /aj/, an initial rise in F1 was ascribed to the preceding consonant, or the transition between the two segments, rather than comprising part of the diphthongal trajectory itself.

1. <u>Demarcation of diphthongs following onset /ɹ/:</u>

In these cases, the first formant F1 and third formant F3 were generally observed to both rise noticeably from the vowel onset to reach a peak or plateau. Low F3 values are typically associated with the articulation of North American English /ɹ/ (see e.g. Hagiwara 1995), which limits the range of F2 as well; in contrast, F1 is not typically associated in any particular way with the articulation of /ɹ/. Vowel onsets were therefore marked at the start of the F1 plateau, indicative of the lowest articulatory position for the incipient nucleus. This point typically coincided with a steep and sudden rise in F3 (see Figure 3.2), although this was not taken as the primary identifier.

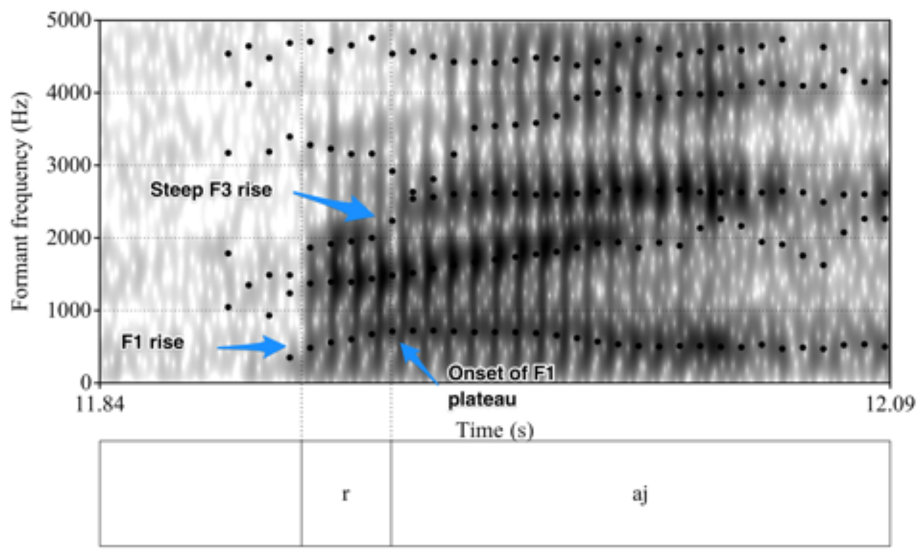


Figure 3.2 Onset /ɹ/: 'ripe', speaker AK69f

2. <u>Demarcation of front off-gliding diphthongs /aj, ɔj/ preceding coda /ɹ/:</u>

In words fitting this pattern, F3 declines (sometimes continuously throughout the vowel) towards and into the /ɹ/. During the off-glide phase of the vowel, F2 reaches a high plateau, shortly after which F3 converges with F2, and both formants thereafter decline in tandem. The vowel offset was marked at this point at which both F2 and F3 begin to decrease simultaneously, indicative of retreat from the furthest forward articulation of the glide portion of the diphthong (see Figure 3.3).

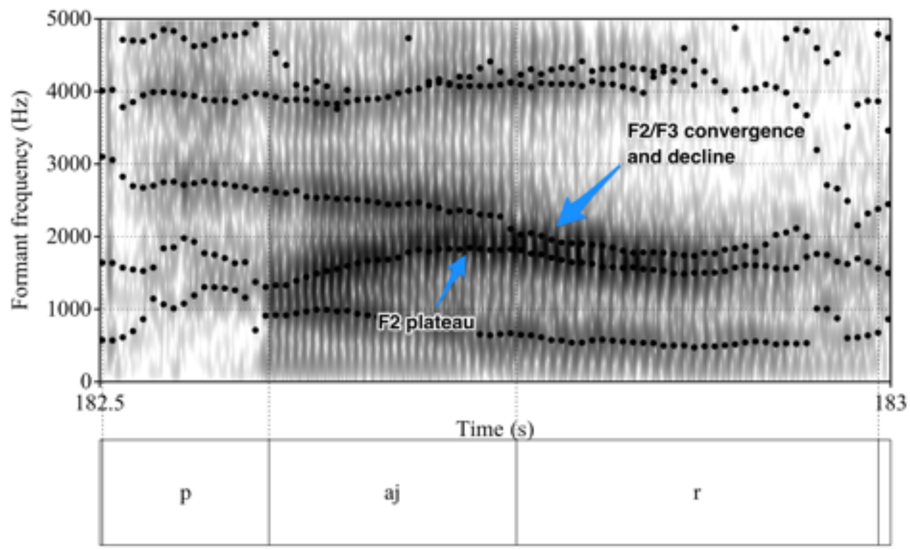


Figure 3.3 Coda /ɹ/ after front glide: 'pyre', speaker AK69f

3. <u>Demarcation of back off-gliding diphthong /aw/ preceding coda /ɹ/:</u>

Here, F2 reaches a low plateau representing the back articulation of the off-glide /w/, after which F3 begins to steadily decline, followed by a rise in F2 as the two formants approach each other and converge. The vowel offset was tagged at the point when F2 begins to rise from its low plateau, indicative of end of the glide articulation, while F3 is typically still declining (see Figure 3.4).

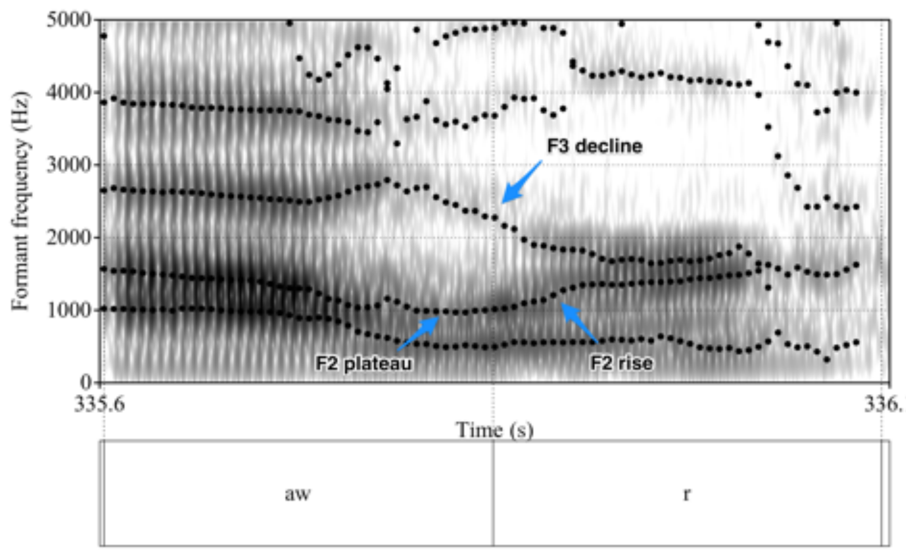


Figure 3.4 Coda /ɹ/ after back glide: 'hour', speaker AK69f

4. Demarcation of front off-gliding diphthongs /aj, ɔj/ preceding coda /l/:

F2 reaches a high plateau during the off-glide phase of the diphthong and then begins to descend, marking the end of the furthest forward position of the glide /j/, at which point F3 ascends at initiation of the articulation of /l/. The vowel offset was marked at the earliest clear point of F2/F3 divergence (see Figure 3.5).

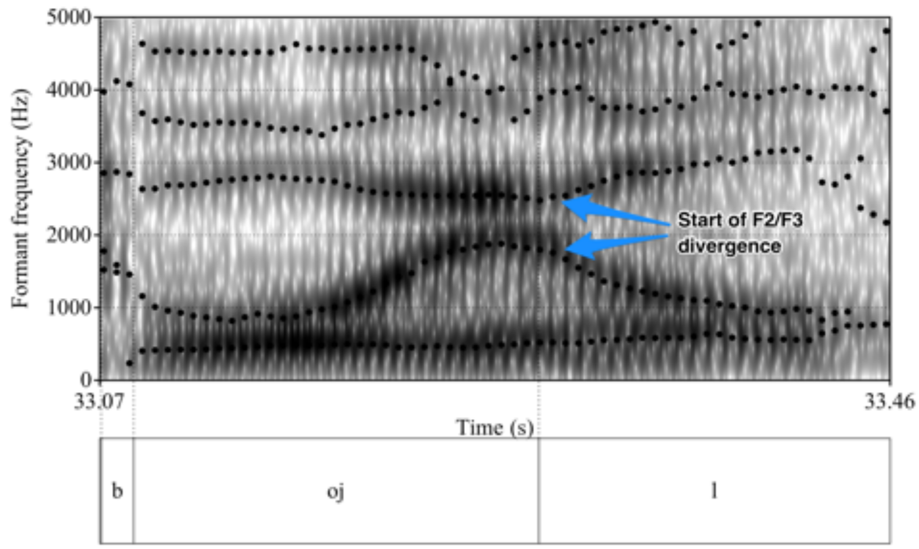


Figure 3.5 Coda /l/ after front glide: 'boil', speaker AK69f

5. <u>Demarcation of back off-gliding diphthong /aw/ preceding coda /l/:</u>

In this case, there is especially little in the way of a clear transition point. This is because /l/ is spectrally very similar to a vowel, and each of the first three formants proceed in distinct, individual trajectories throughout the transition between the segments. F1 and F2 both decline from the nucleus into the glide while F3 rises, each formant continuing on its own trajectory until reaching a plateau, at somewhat staggered (i.e. not entirely coordinated) intervals. The beginning of the F2 plateau was selected as the point to demarcate the /aw/ offset due to its relationship with the back articulation of the glide /w/, as well as the fact that it typically occurred earliest of all the formant plateaus (see Figure 3.6).

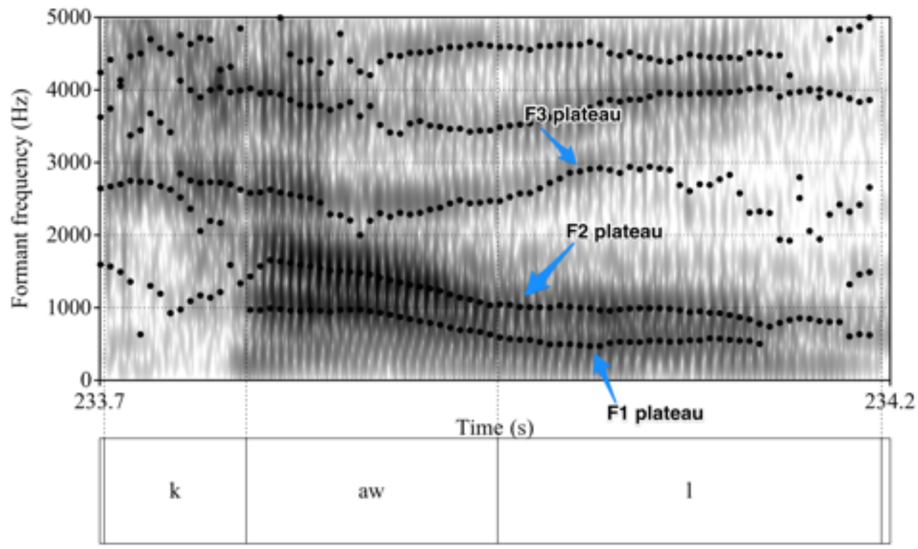


Figure 3.6 Coda /l/ after back glide: 'cowl', speaker AK69f

6. Demarcation of diphthongs preceding coda nasals /m, n/:

There are several notable characteristics which were used to identify the transition from vowel to nasal coda. Spectral intensity is the most easily visible characteristic, observable in the darkness of the formant bands as well as the shape and structure of the waveform. This is aided by the incorporation of the intensity contour track generated by Praat, which shows intensity declining out of the vowel until it typically reaches a plateau of some duration (typically fairly brief) during the nasal. Additionally, simultaneous with decreasing intensity there is a notable change in the periodic nature of the waveform, due to the differing harmonic structures of vowels and nasal consonants. The combined occurrence of lower formant intensity, lower overall intensity, and differing periodic structure was used to demarcate the vowel offset location. Formant intensity was usually the clearest indicator, but as this does not always change uniformly, with the intensity of upper formants (e.g. F3, F4) typically decreasing prior to that of lower formants (F1, F2), a combinatory approach was necessary (see Figure 3.7).

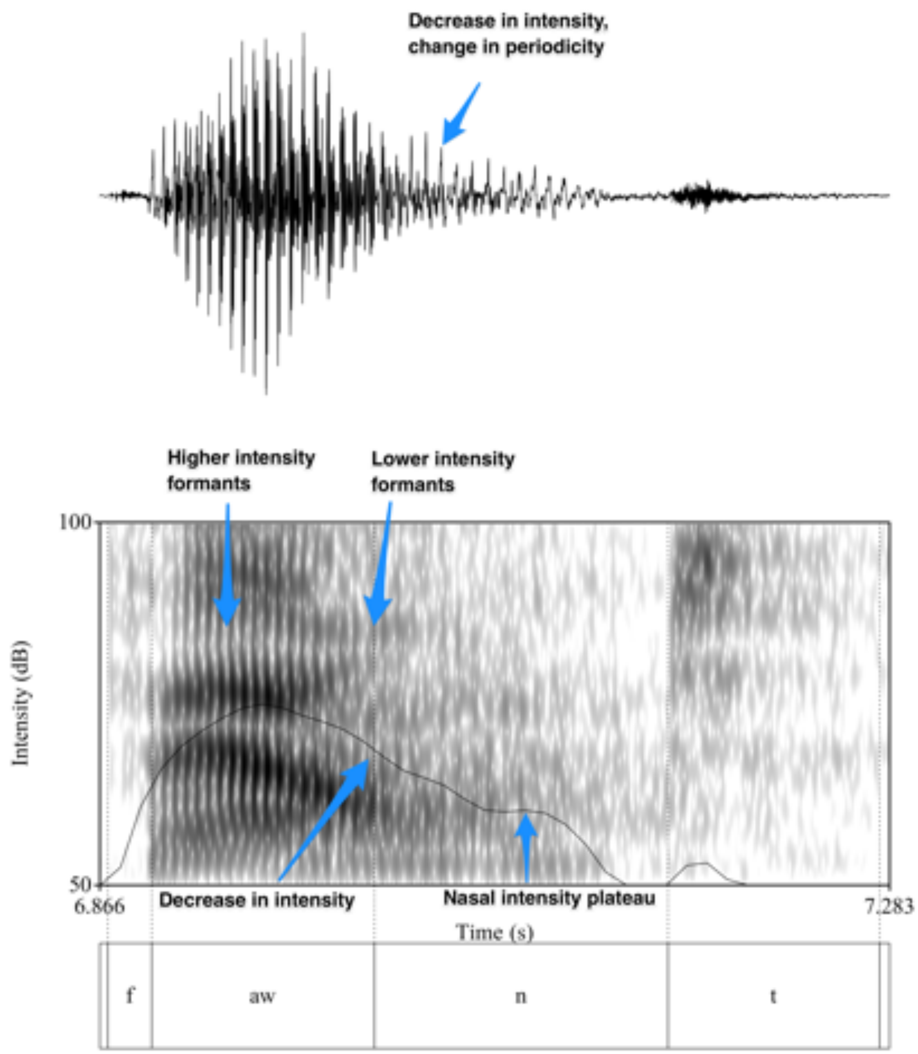


Figure 3.7 Coda nasal: 'fount', speaker AK69f

3.3 VOWEL ANALYSIS

Following the vowel tagging procedures described above, a Praat script (Xu 2013) was run on each audio file to collect a variety of acoustic measurements. For each vowel token, this script captures minimum, maximum, and mean values in Hertz for the first three spectral formants F1, F2 and F3, as well as total vowel duration. Specific formant values for F1, F2 and F3 are additionally measured at 20 evenly-spaced intervals (timepoints) across the vowel duration, at every 5% of the total duration between vowel onset and vowel offset. Hereafter, individual timepoints will be referenced as t1, t2, t3 … t20; the percentile position within the total vowel duration represented by any given timepoint can be calculated by simply multiplying this timepoint numeral by 5%, e.g. t12 = 12 * 5% = 60% of vowel duration. Formant values were unnormalized, and durational data was normalized for certain tests but not others, as described in Chapter 4.

A number of statistical tests were conducted on the resulting data within the R (R Core Team 2016) programming environment running in RStudio (RStudio Team 2016). A small amount of coding was used and modified from existing sources (cited in passing). Scripts and sections of code which were written entirely by me are presented in Appendix H. The results of statistical testing of the data are discussed in detail in the following chapter.

Chapter 4 Results

This chapter discusses the results of statistical analysis of the three datasets collected in Winnipeg, Denver and Madison. Two unfortunate but important omissions must be noted at the outset. As previously mentioned (see Footnote 9), the vowel /ʊ/ before a voiced coda was not elicited for the Winnipeg dataset; the Winnipeg results for /ʊ/ reflect that omission, and /ʊ/ is also therefore excluded from the Denver and Madison data when comparing across cities. Additionally, as the elicitation wordlist was originally designed solely for use in Winnipeg, a dialect with the *cot-caught* merger, no words containing phonemic /ɔ/ were included in the elicitation materials. Although the Madison data represents a dialect (The North) which does not have this merger, i.e. where /ɑ, ɔ/ are distinct phonemes, there are no elicited data for /ɔ/ due to this omission.

This chapter is divided into four subsections, organized as follows. §4.1 describes the acoustic positions of the monophthongal vowel inventory in each sampled city. §4.2 examines patterns of vowel duration in relation to syllable context, specifically coda voicing context, thus providing results pertaining to *pre-voiceless vowel abbreviation* or PVVA. §4.3 describes the positions and dynamic trajectories of the diphthongs in each city. Finally, §4.4 compares diphthong trajectories across different coda-voicing contexts.

4.1 VOWEL POSITIONS

Commonly, sociophonetic results pertaining to overall vowel inventories are reported by plotting mean F1 (vertical) and F2 (horizontal) values, with inverted axes, providing an approximation of articulatory positions within a typical oblong "vowel space". Following this practice, in Figure 4.2 through Figure 4.4, the mean formant values of the vowel inventories (omitting diphthongs) of Winnipeg, Denver and Madison are plotted in such an F1×F2 space, with F2 scaled logarithmically (see discussion in §2.3.3).

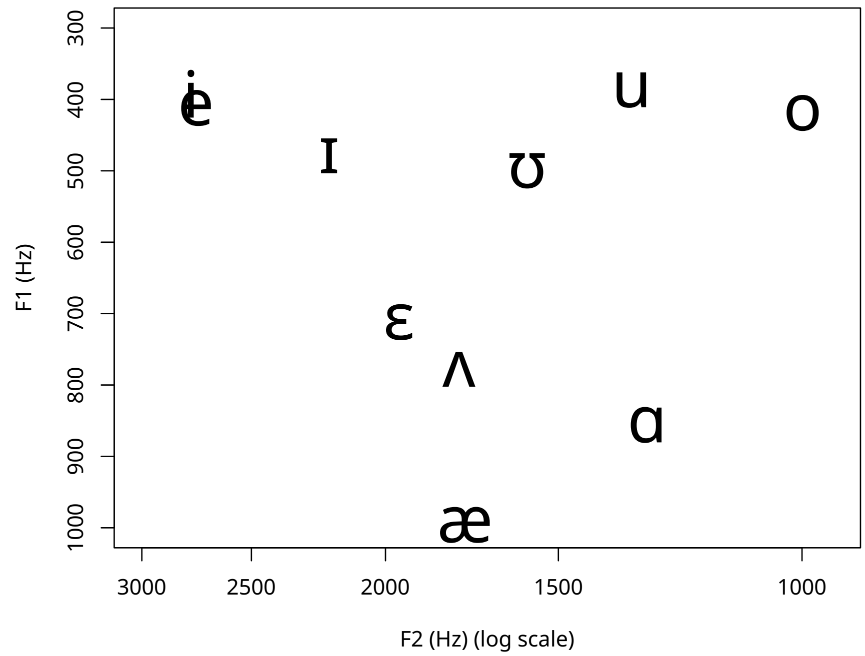


Figure 4.1 Winnipeg women's vowel centres (based on Hagiwara 2006)

The vowel space for female Winnipeggers based on Hagiwara (2006), previously shown in Figure 2.10, is also provided here again as Figure 4.1, allowing direct comparison with the current Winnipeg dataset plot in Figure 4.2[28], below.

---

[28] Although a position is noted for /ʊ/ in Figure 4.2, as this is based solely on pre-voiceless tokens due to erroneously omitting *hood* from the Winnipeg elicitation material, this position is not truly comparable to the indicated position for /ʊ/ in the other datasets, which are based on both pre-voiced and pre-voiceless tokens. Note that in Hagiwara's data in Figure 4.1, /ʊ/ is much higher and advanced in comparison, although there are a number of other differences as well, so it is uncertain whether the same could be expected had pre-voiced tokens been available for the present Winnipeg study.

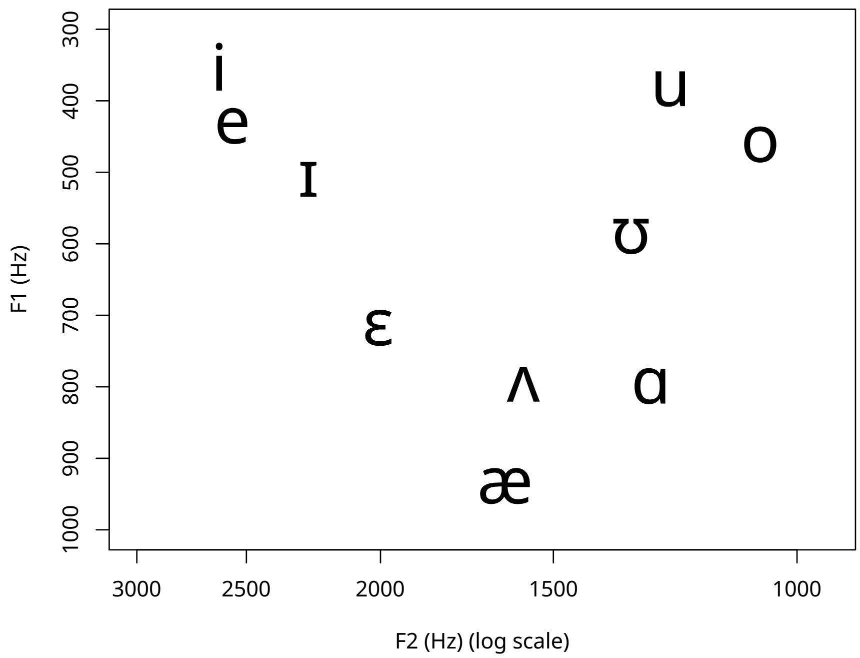


Figure 4.2 Vowels in mean F1–F2 space: Winnipeg

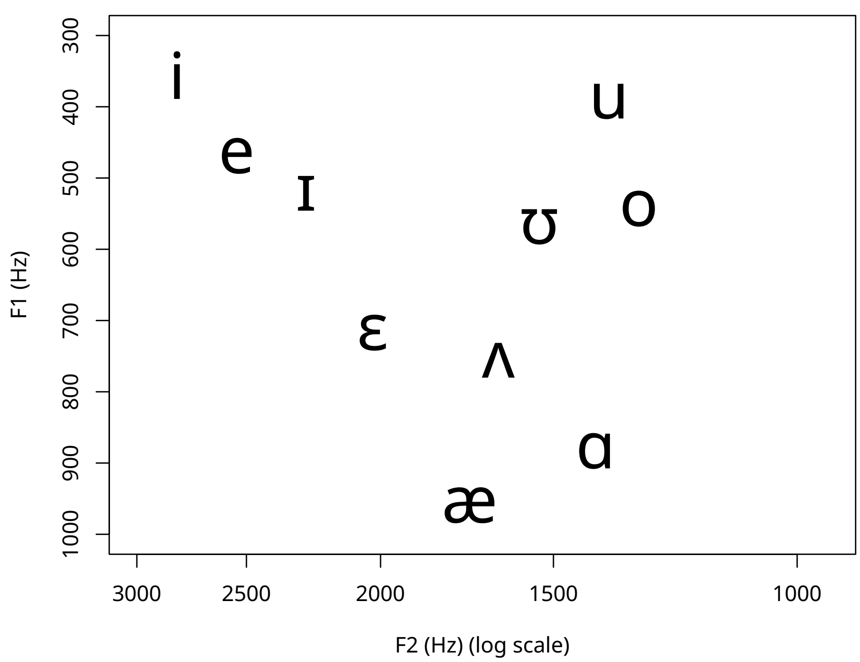


Figure 4.3 Vowels in mean F1–F2 space: Denver

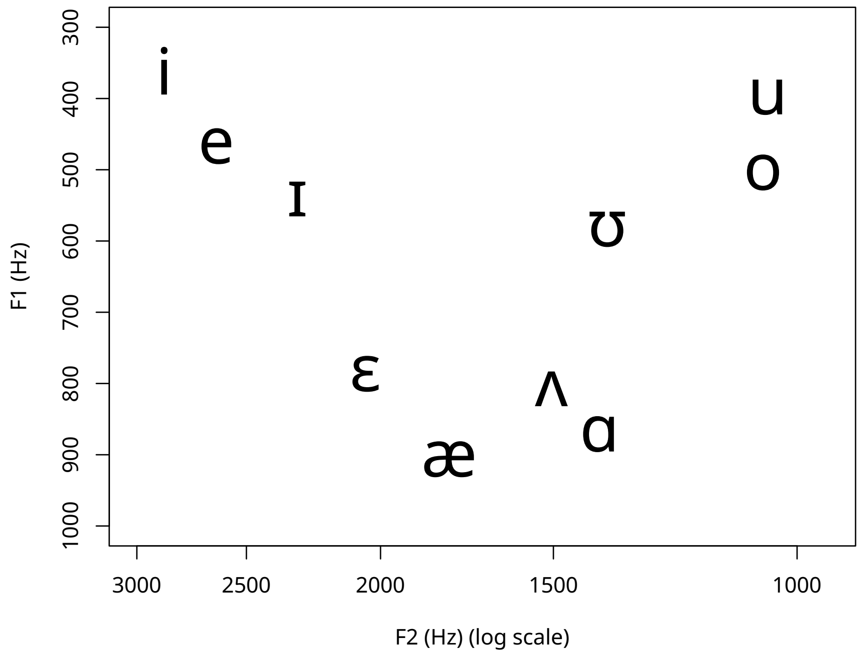


Figure 4.4 Vowels in mean F1–F2 space: Madison

Excluding /ʊ/, I can note the following characteristics of the present Winnipeg dataset in comparison with Hagiwara's results, and with the two American cities:

- /i/ is less advanced than Hagiwara, Denver, or Madison
- /e/ is less advanced than Hagiwara, and higher than both Denver and Madison
- /ɪ/ is higher than both Denver and Madison, i.e. not indicative of Canadian Shift
- /ɛ/ is very slightly retracted compared to Denver and Madison, but higher than Madison; not strongly indicative of Canadian Shift
- /æ/ is more retracted than Hagiwara, as well as both Denver and Madison; strong indicator of incipient Canadian Shift, which typically is initiated by /æ/-retraction following the merger of /ɑ, ɔ/
- /ʌ/ is less advanced and lower than both Hagiwara and Denver, but more advanced than Madison

- /ɑ/ is higher than Hagiwara, Denver and Madison, and less advanced than Denver
- /o/ is advanced in relation to Hagiwara and Madison, but less advanced than Denver
- /u/ is less advanced than Hagiwara or Denver, but more advanced than Madison

Denver exhibits the following notable characteristics:

- /i/ is strongly advanced, more than Winnipeg but similar to Madison
- /e/ is lower than Winnipeg but similar to Madison
- /ʌ/ is both more advanced and more raised than in Winnipeg and Madison
- /u, o/ are more advanced than the other cities; fronting of /u/ is a known feature of The West, and fronting of /o/ is known for Denver specifically (see §2.4.2)
- Relatedly, the entire back portion of the vowel space exhibits a strong degree of overall advancement, with no mean F2 value falling below 1300 Hz, compared with 1063 Hz for Winnipeg /o/, and 1050 Hz for Madison /u/

Madison exhibits the following notable characteristics:

- /ɛ/ is lower than either Denver or Winnipeg, indicative of Northern Cities Shift (see Figure 2.20)
- /æ/ is advanced and raised compared to Denver and Winnipeg, indicative of Northern Cities Shift
- /ɑ/ is advanced and lowered compared to Winnipeg, although similar to Denver, therefore an uncertain indication of Northern Cities Shift; advancement typically *precedes* lowering of /ɛ/, which does occur (see above)
- /ʌ/ is more retracted than Denver or Winnipeg, indicating Northern Cities Shift

- /o/ shows no advancement, being further back than either Denver or Winnipeg, and well below 1200 Hz, as is typical for The North (see §2.5)
- /u/ also shows no advancement, as is typical for The North, aligning closely with /o/ in terms of F2

4.2 VOWEL DURATION PATTERNS

As discussed in §2.1, the abbreviation of vowels in pre-voiceless contexts (PVVA) is well-documented in North American English. It was therefore expected that the three datasets in the present study would all exhibit not only *inherent* durational differences between the various vowels, but also *contextual* durational differences between the allophones of each vowel depending on the presence and voice quality of a following coda consonant.

The distributions of each vowel's duration by syllable type is plotted for each city in Figure 4.5 through Figure 4.7 as beanplots (Kampstra 2008). A beanplot reports a number of useful pieces of statistical data: the total distribution of each vowel's duration, shown via individual token durations (short green horizontal lines); the overall mean for each coda type (medium-length black horizontal lines); and, the mean duration of all tokens for a given vowel (dotted horizontal line spanning an entire plot). Further, the outline of each 'bean' shape indicates the overall distribution pattern for that set of tokens, with wider regions containing a greater quantity of tokens at that duration (higher token density), and thinner regions respectively smaller quantities (lower density). To facilitate cross-study comparisons, all plots follow the same durational scale (y-axis), and results for /ʊ/ are omitted entirely from this section, as they are not available for Winnipeg.

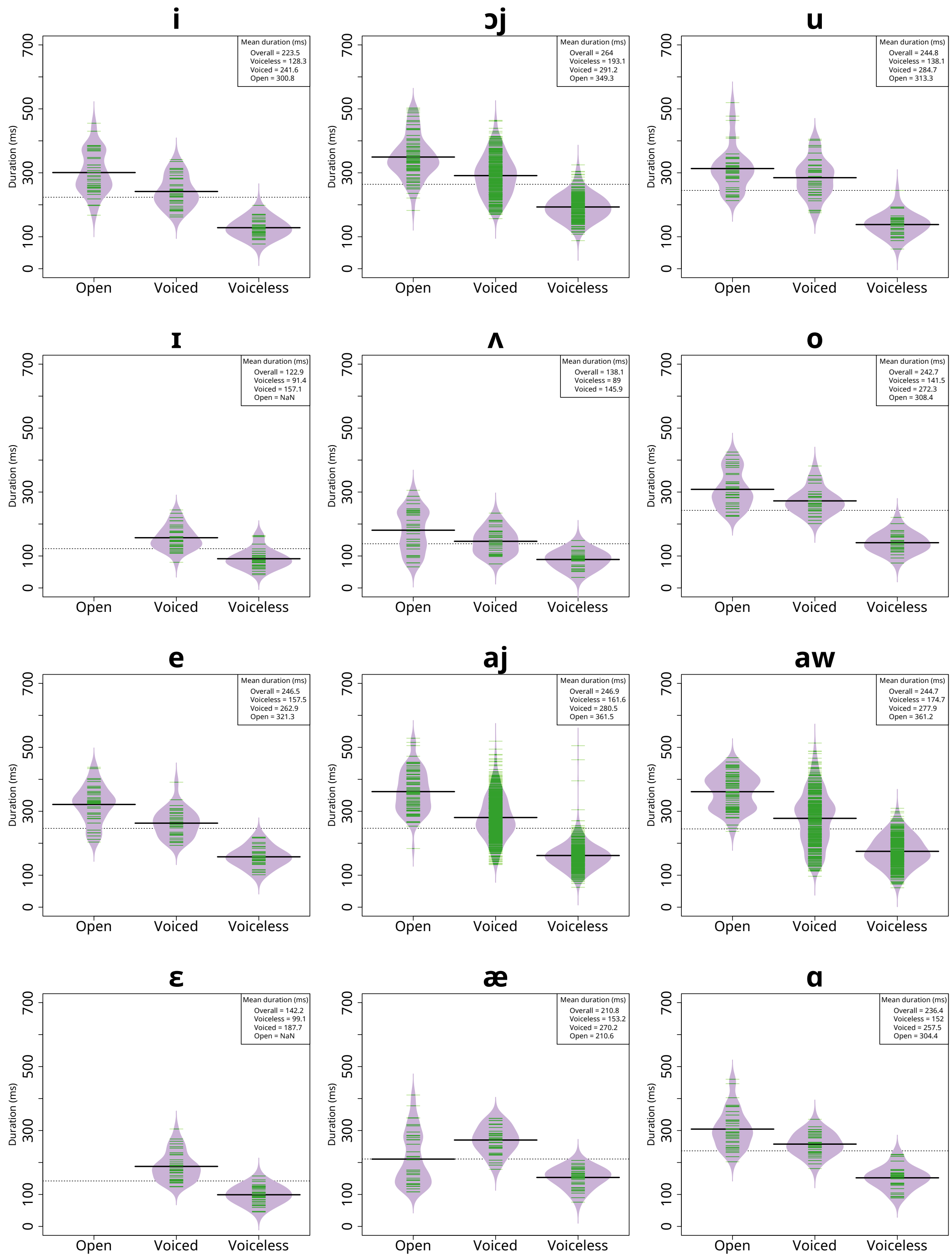


Figure 4.5 Distribution of vowel durations by syllable type: Winnipeg

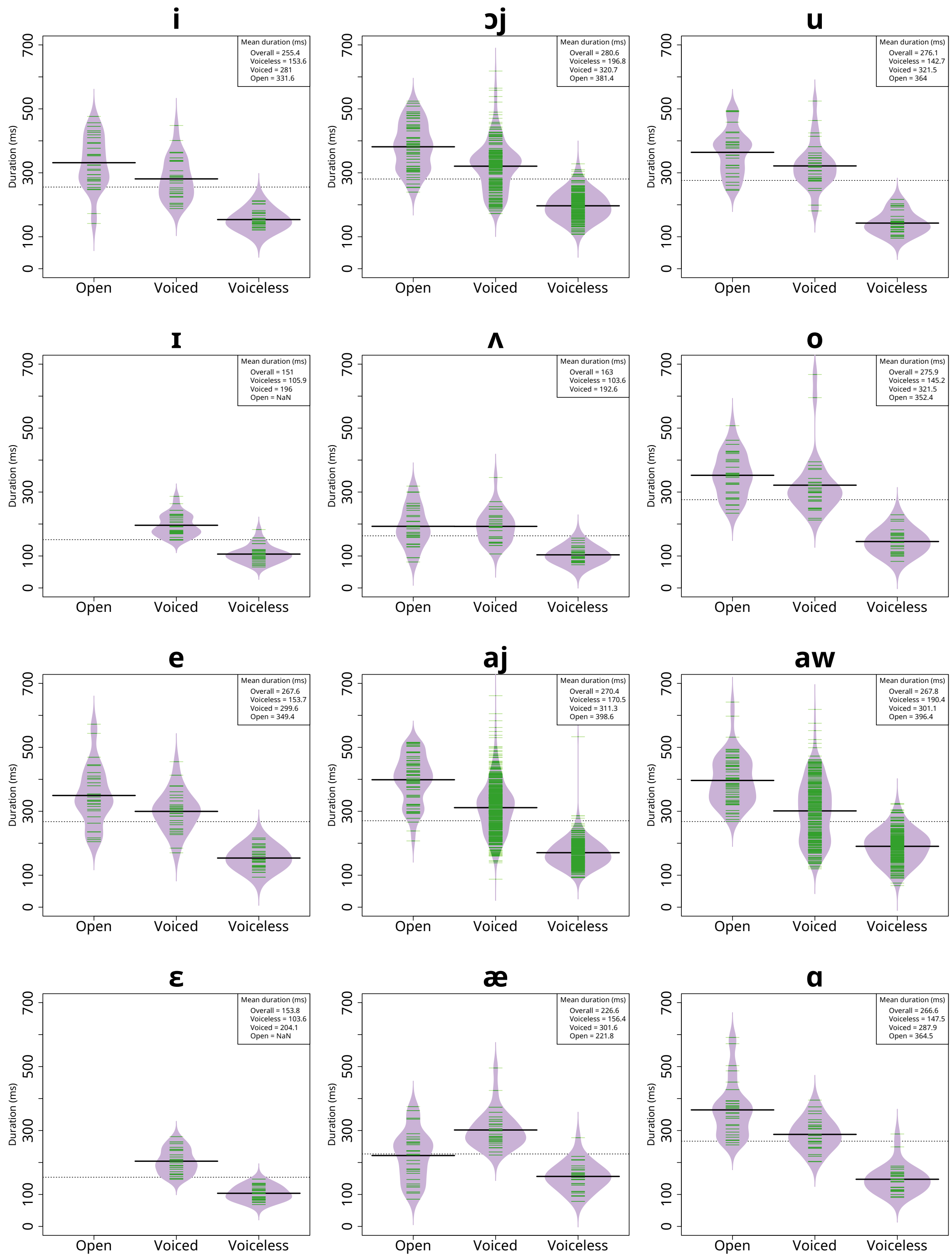


Figure 4.6 Distribution of vowel durations by syllable type: Denver

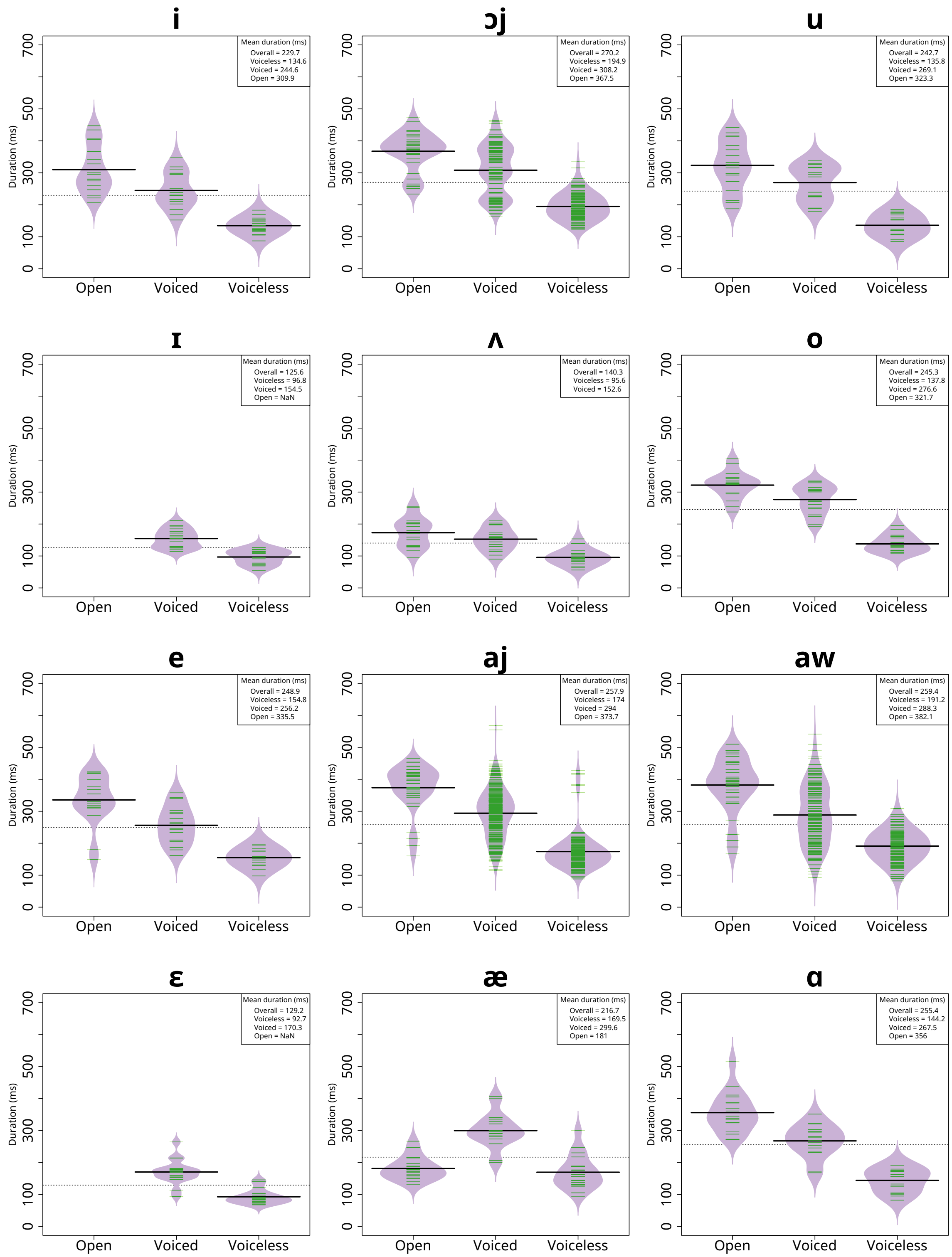


Figure 4.7 Distribution of vowel durations by syllable type: Madison

A cursory examination of the duration distributions reveals that vowels are generally longest in open syllables[29], somewhat shorter before voiced codas, and substantially (and as it turns out, significantly) shorter before voiceless codas. An exception to the generalization that open syllable vowels are longest concerns the vowel /æ/, which has *longer* durations before voiced codas than either in open syllables or before voiceless codas, among speakers in all three cities.

An ANOVA was run within each dataset to test the correlation between the dependent variable of *vowel duration* and the independent variables of *vowel*, with twelve levels (again, excluding /ʊ/), and *syllable type*, with three levels—*open syllable*, *voiced coda*, and *voiceless coda*—as well as the interaction between the independent variables.

Table 4.1 ANOVA of vowel duration by syllable type

	F			
Independent variables	Winnipeg	Denver	Madison	*p*
Syllable type	3029.18	2315.06	1247.21	<0.001
Vowel	86.07	52.66	45.71	<0.001
Interaction: *Syllable type* & *vowel*	14.39	10.69	8.23	<0.001
Token quantity, N =	3,068	2,389	1,430	

As shown in Table 4.1, each independent variable, and the interaction between the two, was found to be highly significant (p<0.001) within each dataset; due to the significant interaction, post hoc tests were also conducted for each dataset (see below). Effect sizes (F) vary from city to city, but are consistent in terms of the relative effect sizes for each variable: *syllable type* is by far the largest effect across the board, with the effect of *vowel*

[29] Note that the lax vowels /ɛ, ɪ/ do not occur in open syllables in English, so those distributions are absent from their respective plots.

being substantially smaller, and the interaction between the two being relatively minor. The differences in relative effect sizes between cities are most likely at least partially related to the size of each dataset. This can be seen by comparing the token quantity of each dataset as reported in the last line of the table; relative to the Winnipeg dataset, Denver contains approximately 78% as many tokens, and Madison only 47% as many.

Table 4.2 Cross-dialect comparison of durational differences between syllable types

	Mean duration difference (ms)		
Syllable types compared	Winnipeg	Denver	Madison
Open syllable – voiced coda	43.96	48.44	46.43
Open syllable – voiceless coda	154.77	179.28	158.18
Voiced coda – voiceless coda	110.81	130.85	111.75

Post hoc tests covered three sets of comparisons: *syllable types*, individual *vowels*, and the combination of *syllable type* and *vowel*. For the first of these, within each vowel inventory, each of the three *syllable types* was found to differ significantly from the other two with respect to the effect on *duration*, as summarized in Table 4.2; in every case, the differences were of the highest significance, with $p=0$ in every case. Overall, Denver reports the largest mean differences between syllable types, and Winnipeg the smallest, with Madison falling medially between the other two cities.

The next set of post hoc comparisons revealed those vowels which are significantly different from each other in terms of their overall duration (non-subdivided for syllable context), which varies between cities. At a confidence measure of $p<0.001$ (a high level of significance), Winnipeg reports 39 of 66 available pairwise vowel comparisons, or 59.9% are significantly different in terms of overall duration; Denver reports 31 of 66, or 46.9%; and Madison reports 33 of 66, or 50%, being once again medial to the other two cities.

Comparison of the inherent vowel duration results across the three datasets reveals that areas of difference between them concern only three vowels: /ɔj, æ, i/. Comparing Winnipeg to Denver, nearly all of the differences concern /ɔj/. In Winnipeg, the duration of /ɔj/ is significantly different from each of /ɑ, e, o, u, aj/, none of which are significantly different from /ɔj/ in Denver. Winnipeg also exhibits significant durational differences between /æ/ vs. /e, ɪ/, and between /aj/ vs. /i/; these particular pairs are non-significantly different in Denver. Finally, in Denver /æ/ vs. /o/ are significantly different, which is not the case for Winnipeg. Madison again appears to occupy a medial position with regard to inherent vowel duration differences. For /ɔj/, Madison exhibits more significant differences than Denver—adding the vowels /o, u/—but fewer than Winnipeg—lacking any significant difference from the vowels /ɑ, e, aj/. Other differences in Madison centre mainly around the vowel /æ/, which has fewer durational differences from other vowels than the other two cities, lacking difference from the durations of /e, ɪ, u/ (from which it is distinct in Winnipeg) and /o/ (Denver). And finally, Madison exhibits a significant difference between /i/ and both /aj, aw/, in contrast with Winnipeg which lacks a significant difference between /i/ and /aw/, and with Denver where /i/ differs from neither diphthong.

Table 4.3 Significantly different inherent vowel durations, three cities compared

City	/æ/ vs.	/i/ vs.	/ɔj/ vs.
Winnipeg	ʌ … i aw u e aj ɔj	ɪ ʌ ɛ … aj ɔj	ɪ ʌ ɛ æ i ɑ o aw u e aj …
Denver	ʌ … aw aj o u ɔj	ɪ ɛ ʌ … ɔj	ɪ ɛ ʌ æ i …
Madison	ʌ … aj aw ɔj	ɪ ɛ ʌ … aj aw ɔj	ɪ ɛ ʌ æ i u o …

Table 4.3 summarizes all of the differences between the three cities in pairwise comparisons of significantly different vowel durations. For visual comparison purposes, mean vowel durations for each city are plotted in Figure 4.8 through Figure 4.10, ordered

from least to greatest mean duration. The ordering of vowels in Table 4.3 reflects the per-city visual order as presented in the respective figure for that city.

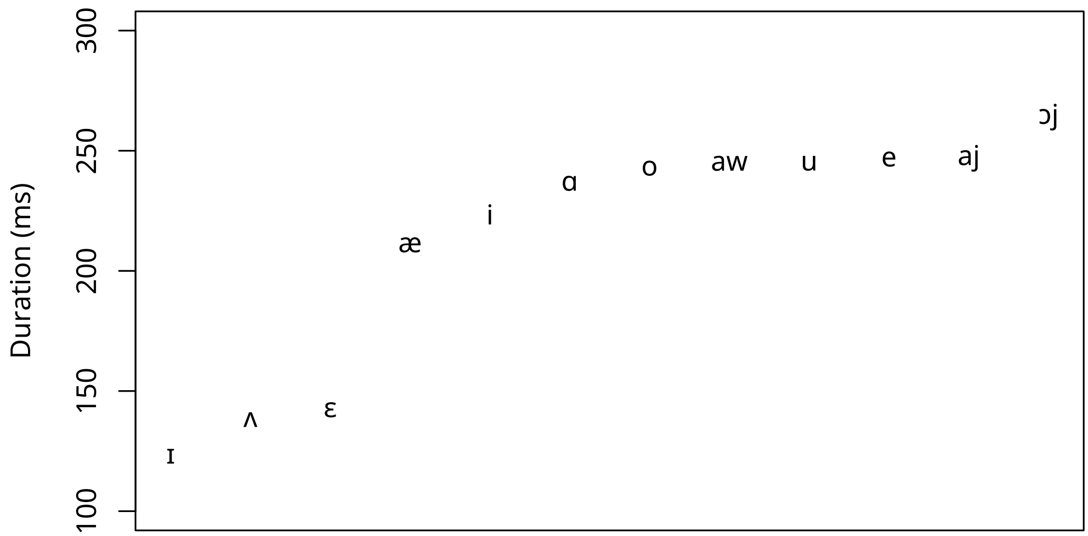


Figure 4.8 Mean vowel durations: Winnipeg

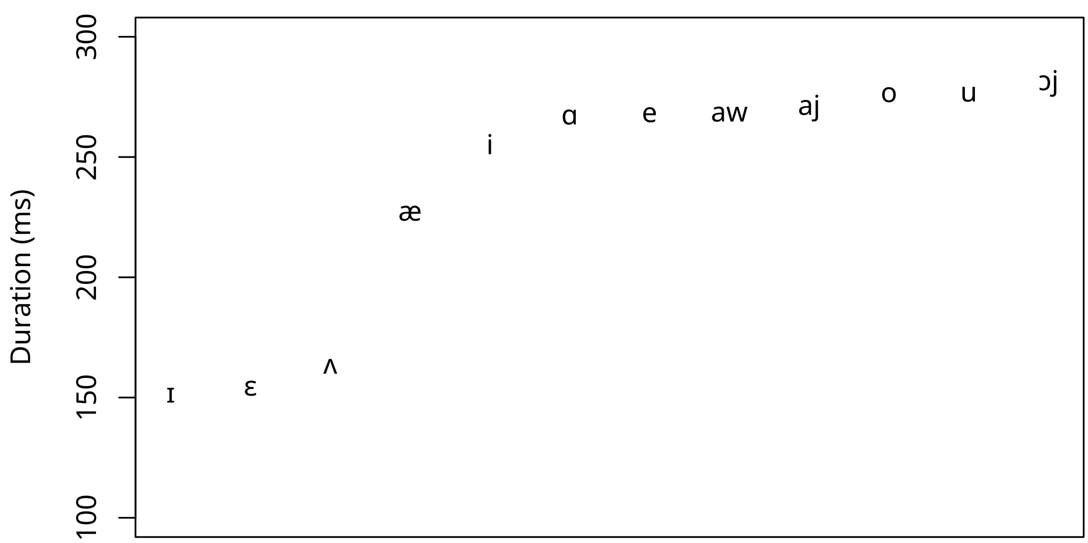


Figure 4.9 Mean vowel durations: Denver

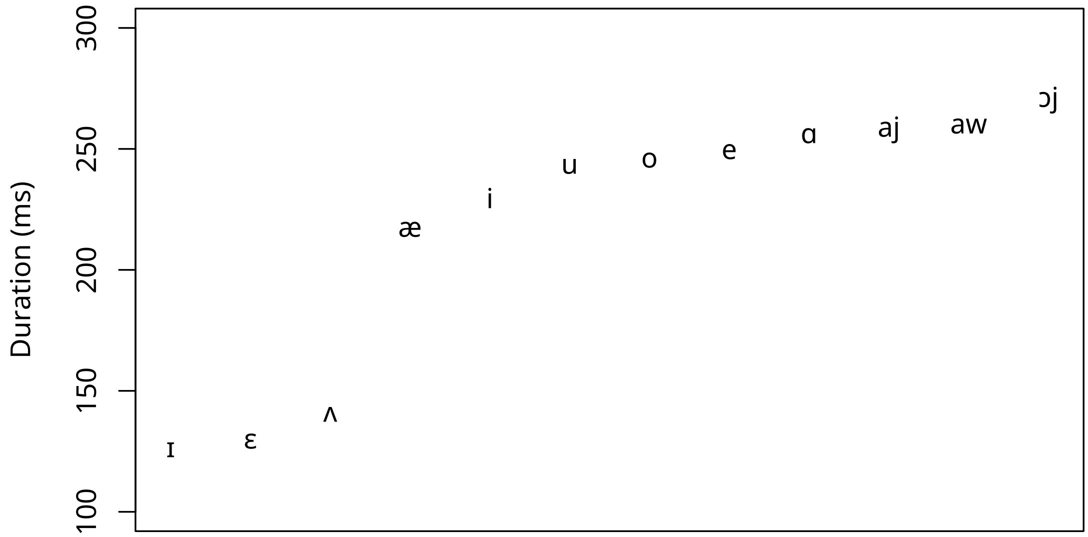


Figure 4.10 Mean vowel durations: Madison

Finally, post hoc tests were run for the combination of *syllable type* and *vowel*. The majority of syllable-vowel comparisons had significantly different durations ($p<0.05$), but there were substantial differences between cities in terms of those combinations whose durations did not differ significantly ($p>0.05$). Madison had the largest number of non-significantly different pairs, and was the only dataset which contained vowel durations which did not differ significantly between voiced and voiceless codas; these were all lax vowels, and included: /ɛ, ɪ, ʌ/. In addition, Madison had a large number of vowels whose durations in open syllables and before voiced codas did not differ significantly: /æ, e, i, o, ʌ, u/. Finally, for /ʌ/ there was also no significant difference in duration between open syllables and voiceless codas; this was the only case of a vowel which exhibited no significant durational difference across all syllable types. Recall (§2.5.1) that the region including Madison has been argued to exhibit some Germanic influence with respect to the coda voicing distinction, which can include expression of the PVVA pattern; it may be that these results reflect such influence.

Winnipeg and Denver exhibited far fewer non-significant differences across syllable types for individual vowels. In Winnipeg, such non-significant differences were restricted to the combination of open syllables and voiced codas, for the vowels /ɑ, o, ʌ, u/. In Denver, the same pattern applied to the vowels /e, i, o/; in addition, for the vowel /æ/ in Denver there was no significant durational difference between open syllables and before voiceless codas.

Table 4.4 Non-significant (*p*>0.05) durational differences between syllable types

Syllable types	Winnipeg	Denver	Madison
Voiced and voiceless coda	*none*	*none*	ɛ, ɪ, ʌ
Open syllable and voiced coda	ɑ, o, ʌ, u	e, i, o	æ, e, i, o, ʌ, u
Open syllable and voiceless coda	*none*	æ	ʌ

All of the above-noted patterns of *non*-significant durational differences are summarized in Table 4.4. All syllable type comparisons across vowels absent in Table 4.4 were found to (positively) exhibit significantly different durations. Comparing across the per-city results, there are few observable patterns. Madison is an outlier in terms of both the quantity and scope of its non-significant durational differences. Winnipeg and Denver are not especially similar to each other, however, aside from the fact that in both cities all vowels exhibit significant durational differences between voiced and voiceless codas. The single commonality across all three cities concerns the vowel /o/, whose duration is uniformly non-distinct between open syllables and voiced codas. In addition to this, Winnipeg and Madison are similar in terms of the patterning of the vowels /ʌ, u/ across open syllables and voiced codas; and Denver and Madison are similar in terms of the patterning of /e, i/ in the same contexts.

Overall, the results of ANOVA testing confirm the general occurrence of PVVA in all three represented dialects—Canada, The West, and The North—with the caveat that Madison exhibits some non-PVVA patterning among lax vowels. PVVA can be quantified as the ratio of mean vowel duration in pre-voiceless vs. pre-voiced contexts, for individual vowels as well as across the entire vowel system. This *duration ratio* has an inverse relationship with PVVA. A duration ratio of 1.0 would indicate that pre-voiceless durations are equal to pre-voiced durations, i.e. PVVA does not occur. Increasingly smaller ratios indicate an inversely increasing degree of PVVA, e.g. a ratio of 0.6 indicates *more* abbreviation in pre-voiceless context than does a ratio of 0.7.

Studies which have examined the relationship between vowel duration and coda voicing in American English dialects have reported results which indicate (overall) PVVA duration ratios ranging from 0.548 (House 1961) to 0.721 (Whitehead & Jones 1976) for speakers with typical hearing and speech. Table 4.5 provides a comparison of mean vowel duration ratios across the various studies cited in §2.1 for which duration ratios are calculable, including the comparative studies involving deaf, hearing-impaired, and esophageal speakers, along with the results from the three studies carried out for this dissertation, which are cited as Onosson (2018). It should be noted that the same set of vowels were not compared across studies, with some examining only a small subset of the vowel inventory, often comprising just two to four vowels. Studies which are broadly comprehensive in terms of the vowel range surveyed are especially rare among older studies, with Peterson & Lehiste (1960) being a notable exception. I omit Tauberer & Evanini (2009) which, despite being one of the most recent and comprehensive studies, obtained a range of highly divergent (larger) ratios as compared with other results, likely due to methodological differences (see discussion at the end of §2.1.1). Hall's (2016b) results are presented separately for each of the two CR diphthongs, in each of the two sampled cities in that study. For the present study of Madison, I omit the three lax vowels which do not exhibit PVVA-patterning, namely /ɛ, ɪ, ʌ/, when calculating that city's duration ratio.

Table 4.5 Ratio of vowel durations by coda voicing across multiple studies

Study	Dialect	Vowel(s)	Dur. ratio
Whitehead & Jones (1976)	American English*; *deaf*	/i, ɑ/	0.853
Pycha & Dahan (2016)	American English; *multiple dialects*	/aj/	0.792
Whitehead & Jones (1976)	American English*; *hearing-impaired*	/i, ɑ/	0.768
Whitehead & Jones (1976)	American English*, *no hearing deficit*	/i, ɑ/	0.721
Hall (2016b)	Canadian English; *Toronto*	/aw/	0.709
Luce & Charles-Luce (1985)	American English*	/i, ɪ, ɑ/	0.69
House & Fairbanks (1953)	American English*	/i, e, æ, ɑ, o, u/	0.688
Hall (2016b)	Canadian English; *Vancouver*	/aw/	0.667
Klatt (1973)	American English*	*unspecified*	0.667
Peterson & Lehiste (1960)	American English*	*full inventory*	0.663
Sharf (1964)	American English*	*monophthongs*	0.656
Hall (2016b)	Canadian English; *Vancouver*	/aj/	0.638
Gandour et al. (1980)	American English*; *laryngeal speech*	/i, ɑ, u/	0.633
Sharf (1964)	American English*; *whispered*	*monophthongs*	0.62
Hall (2016b)	Canadian English; *Toronto*	/aj/	0.613
Gandour et al. (1980)	American English*; *esophageal speech*	/i, ɑ, u/	0.574
Onosson (2018)	**Canadian English; *Winnipeg***	*full inventory*	**0.573**
Onosson (2018)	**American English; *Madison***	*full inventory*	**0.572**
House (1961)	American English*	*monophthongs*	0.548
Onosson (2010)	Canadian English; *Winnipeg*	/aj/	0.539
Onosson (2018)	**American English; *Denver***	*full inventory*	**0.53**

* *Dialect unspecified*
Results from present study (Onosson 2018) in boldface

There are a wide range of ratios observed in Table 4.5, with the various American English (no hearing deficit) studies ranging from 0.721 (Whitehead & Jones 1976) to 0.53 (the present Denver results). Overall, Canadian studies, though largely limited to just the

CR diphthongs, cover a similar if somewhat smaller range, from 0.539 (Winnipeg /aj/; Onosson 2010) to 0.709 (Toronto /aw/; Hall 2016b); the sole Canadian study to include the full vowel inventory, namely the Winnipeg data collected for this dissertation, is towards the low end of the PVVA spectrum at 0.573.

It seems reasonable to speculate that the wide range of ratios found amongst the cited studies may be at least partially ascribable to methodological differences such as the range of vowels included in each study, which as noted varies considerably, or segmentation protocols (see §3.2). Focusing on the findings from only the present three studies, which share methodologies and are therefore directly comparable, the ratios for Winnipeg and Madison are nearly identical at 0.573 and 0.572, respectively, while Denver has a much smaller ratio of 0.53, indicating that PVVA has a substantially larger effect on vowel durations there than in the other two cities.

After establishing the mean PVVA ratios across all vowels per city, voiceless-to-voiced coda duration ratios were calculated for each individual vowel, as shown in Table 4.6 through Table 4.8, ordered by duration ratio from largest to smallest, i.e. least to greatest PVVA; again, for Madison those vowels with non-significant durational differences between voiced and voiceless codas were omitted.

Table 4.6 Vowel duration by coda voice context: Winnipeg

Vowel	Duration		Duration ratio
	Voiceless coda	Voiced coda	
ɔj	193.09 ms	291.19 ms	0.663
aw	174.66 ms	277.87 ms	0.629
ʌ	89.014 ms	145.9 ms	0.61
e	157.55 ms	262.86 ms	0.599
ɑ	152.03 ms	257.49 ms	0.59
ɪ	91.37 ms	157.07 ms	0.582
aj	161.61 ms	280.46 ms	0.577
Mean	*139.96 ms*	*244.11 ms*	*0.573*
æ	153.18 ms	270.2 ms	0.567
i	128.27 ms	241.57 ms	0.531
ɛ	99.06 ms	187.69 ms	0.528
o	141.52 ms	272.31 ms	0.52
u	138.13 ms	284.67 ms	0.485

Table 4.7 Vowel duration by coda voice context: Denver

Vowel	Duration		Duration ratio
	Voiceless coda	Voiced coda	
aw	190.38 ms	301.13 ms	0.632
ɔj	196.78 ms	320.68	0.614
aj	170.49 ms	311.32 ms	0.548
i	153.64 ms	280.98 ms	0.547
ɪ	105.92 ms	196.03 ms	0.54
ʌ	103.6 ms	192.58 ms	0.538
Mean	*147.5 ms*	*278.24 ms*	*0.53*
æ	156.4 ms	301.63 ms	0.519
e	153.72 ms	299.63 ms	0.513
ɑ	147.5 ms	287.85 ms	0.512
ɛ	103.64 ms	204.05 ms	0.508
o	145.19 ms	321.45 ms	0.452
u	142.68 ms	321.53 ms	0.444

Table 4.8 Vowel duration by coda voice context: Madison

	Duration		
Vowel	Voiceless coda	Voiced coda	Duration ratio
aw	191.21 ms	288.26 ms	0.663
ɔj	194.88 ms	308.17	0.632
e	154.83 ms	256.23 ms	0.604
aj	174.02 ms	294.01 ms	0.592
Mean	*159.66 ms*	*278.23 ms*	*0.572*
æ	169.54	299.61 ms	0.566
i	134.64 ms	244.59 ms	0.55
ɑ	144.21 ms	261.57 ms	0.539
u	135.82 ms	269.11 ms	0.505
o	137.76 ms	276.58 ms	0.498
ɪ	96.77 ms	154.5 ms	n.s.d.
ʌ	95.58 ms	152.59 ms	n.s.d.
ɛ	92.69 ms	170.25 ms	n.s.d.

n.s.d. = *no significant difference*

As a means of visually comparing individual vowel PVVA ratios, plots showing the intersection of each vowel's pre-voiceless and pre-voiced duration are provided in Figure 4.11 through Figure 4.13, excluding the lax vowels /ɪ, ʌ, ɛ/ which had non-significantly different durations across coda contexts in Madison. The solid red line crossing each plot indicates the mean duration ratio for that dataset; vowels falling above this line have a larger than average ratio, i.e. a lesser degree of PPVA, and those below the line have a smaller ratio, i.e. a greater degree of PVVA. Dashed lines indicate the deviation of each vowel's mean PVVA ratio from the PVVA ratio for the overall vowel inventory. The y-axis value where the dashed line meets the mean PVVA ratio line indicates the expected pre-voiceless duration for that vowel based on the overall mean across the vowel inventory; the y-axis value at the position of the vowel symbol indicates the observed mean pre-voiceless duration for that vowel.

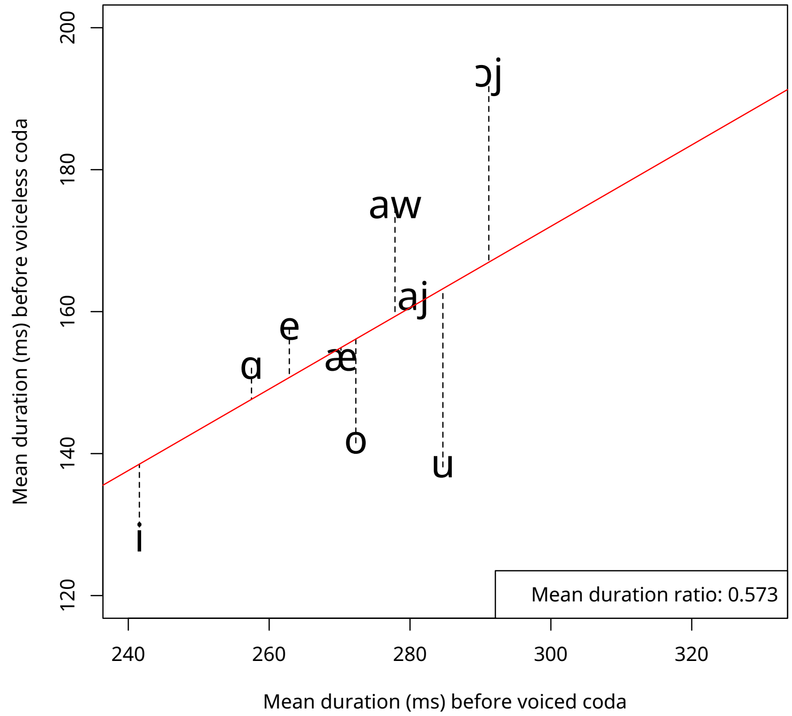


Figure 4.11 Mean vowel duration by coda voice: Winnipeg

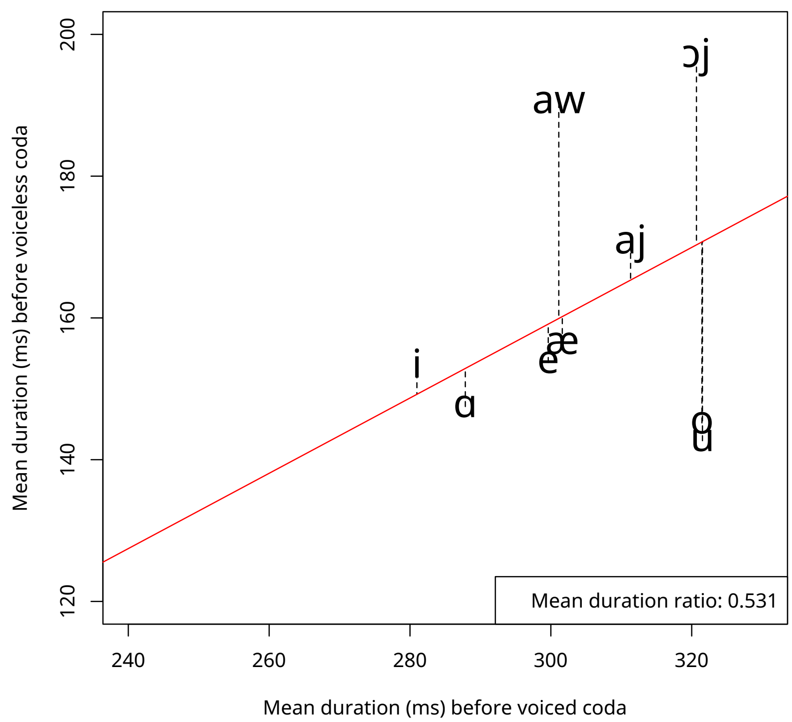


Figure 4.12 Mean vowel duration by coda voice: Denver

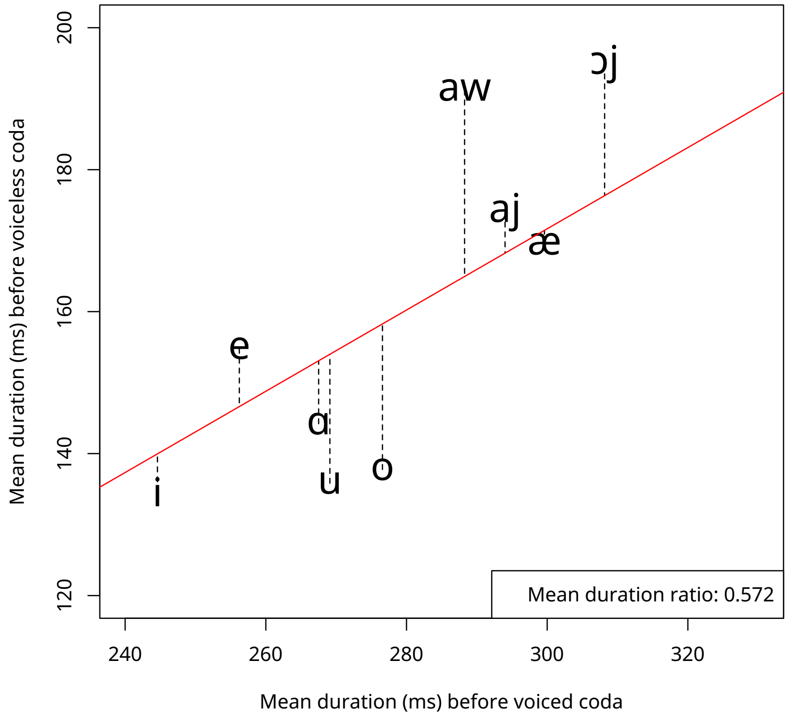


Figure 4.13 Mean vowel duration by coda voice: Madison

Figure 4.14 illustrates the different PVVA ratios obtained for the three diphthongs in each dataset, combined in a single chart; the mean ratios in this chart differ from those indicated in the previous figures, as they are calculated from the diphthongal vowel inventory only.

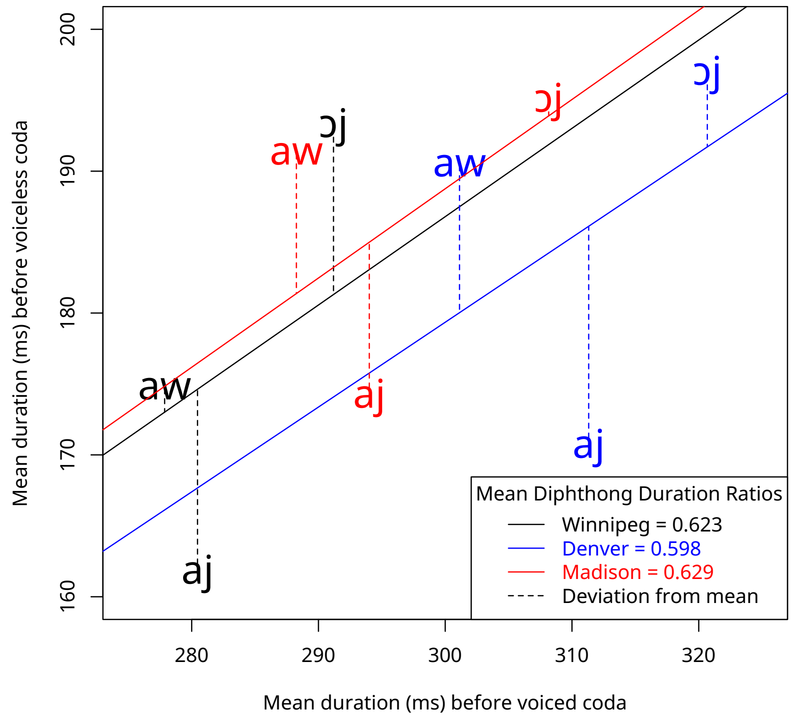


Figure 4.14 Mean vowel duration by coda voice and PVVA ratios, diphthongs only

The most substantial cross-dialectal difference among the diphthongal PVVA patterns concerns the relative positions of /aw, ɔj/. For Denver and Madison, /aw/ deviates the furthest above the mean PVVA ratio, such that it exhibits the least degree of PVVA, while in Winnipeg /aw/ sits fairly close to the mean. Conversely, in Denver and Madison, /ɔj/ sits near or somewhat above the mean, while in Winnipeg it substantially exceeds the mean. In all three cities, /aj/ falls well below the mean diphthong PVVA ratio.

The findings reported in this section establish that PVVA plays a significant role in vowel production in each of the three cities examined, but that there are also important differences between the three dialects represented by those populations. In Denver, mean duration differences between syllable types, including between voiced and voiceless codas i.e. PVVA itself, are greatest among the three cities, while in Winnipeg these same durational differences are the smallest, while Madison falls in between. When it comes to inherent vowel duration differences, in pairwise comparisons between individual vowel durations, this order is reversed with Winnipeg exhibiting the largest number of significant durational differences and Denver the smallest, with Madison again falling in between. When comparing PVVA ratios by vowel, in each city the diphthongs /aw, ɔj/ consistently had the largest ratios (least PVVA) and the vowels /o, u/ the smallest ratios (most PVVA). When comparing PVVA ratios exclusively among the diphthongs, the relative positions of /aw, ɔj/ are reversed between Denver and Madison on the one hand, where /aw/ exhibits the largest ratio, and Winnipeg on the other hand, where that status applies to /ɔj/. Finally, the relatively low PVVA ratio of /aj/ compared to the other diphthongs is similar across all three dialects, and consistently close to the overall mean PVVA ratio in each city.

4.3 DIPHTHONG POSITIONS AND TRAJECTORIES

In §4.1 vowel positions in each dialect were examined, excluding the diphthongs /aj, aw, ɔj/. Plotting static positions of diphthongs (e.g. Figure 2.8) necessarily obscures their dynamic characteristics. The use of multiple measurement points (i.e. at least two) permits observation of at least some aspect of the dynamic, changing trajectories typical of diphthongs. This method was used by Hagiwara (2006) who recorded the values of F1 and F2 for each diphthong at three timepoints, at 25%, 50% and 75% of vowel duration.

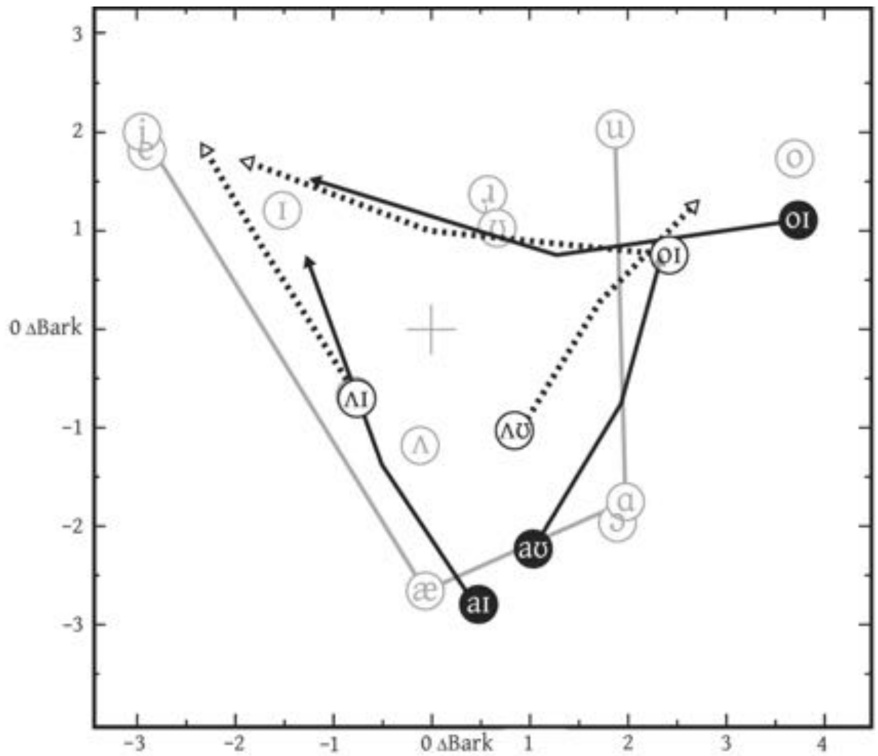


Figure 4.15 Women's diphthongs: Winnipeg (Hagiwara 2006:137)

Figure 4.15 plots Hagiwara's diphthong trajectories in relation to coda voice; pre-voiced allophones are indicated by dark-shaded circles at the initial articulatory position (actually 25%, the closest available proxy) with solid lines showing the path of the trajectory, and pre-voiceless allophones are indicated by unshaded circles with dotted lines, and the monophthongal vowel inventory is also shown in lightly shaded text in the background of the figure. Although the use of multiple (in this case, three) timepoints reveals a fair amount of information about the plotted trajectories, the limited number of timepoints used in Figure 4.15 obscures a few important details of diphthong articulations, which are revealed when substantially more measurement points are included; this was the method used for the data from the cities investigated in the present study.

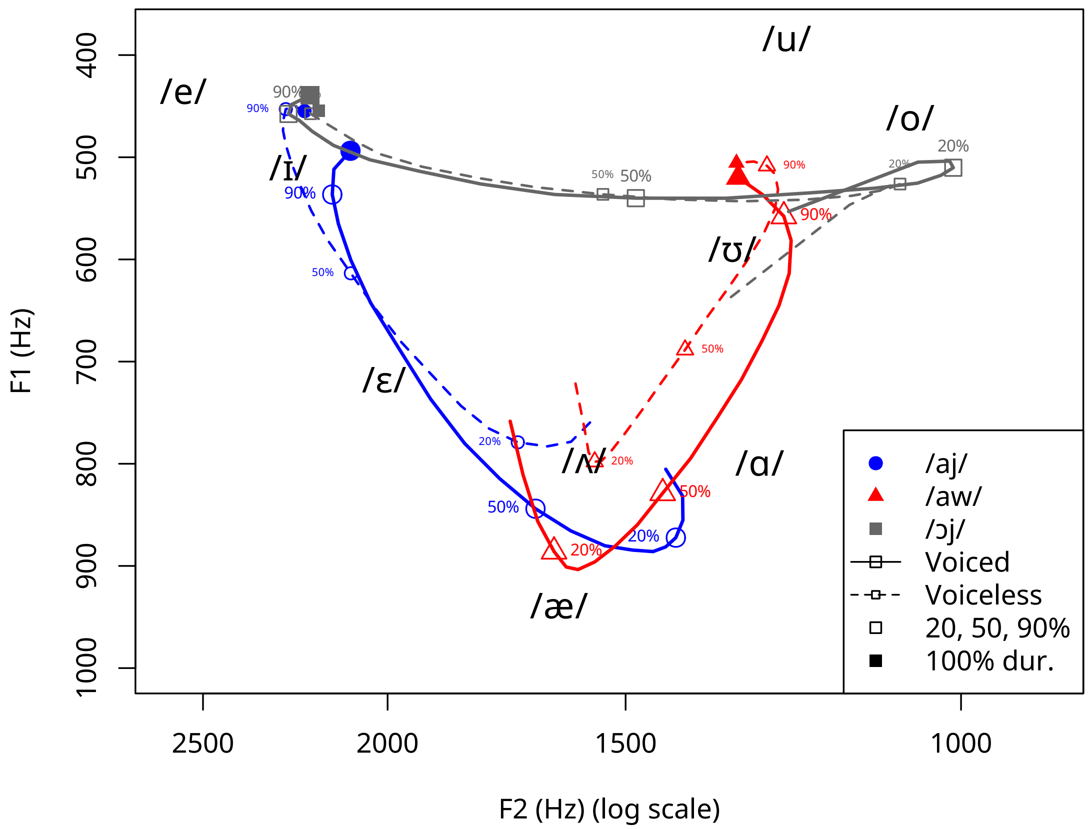


Figure 4.16 Diphthong trajectories: Winnipeg

The diphthong trajectories from the present Winnipeg dataset are shown in Figure 4.16, utilizing formant measurements at each of 20 timepoints across the duration of each token. Each diphthong's trajectory is indicated by a particular colour and symbol, as shown in the legend: blue circles for /aj/; red triangles for /aw/; and gray squares for /ɔj/. Pre-voiced trajectories are indicated by solid lines and larger symbols, pre-voiceless trajectories with dashed lines and smaller-sized symbols. Each trajectory follows the mean F1×F2 values taken at 5% duration intervals (see §3.3). Open, unshaded symbols mark several intermediate positions along the trajectory: 20%, 50% and 90% of vowel duration; 50% is an arbitrary value which simply indicates the "halfway point" of each trajectory, but the choices of 20% and 90% are more motivated, as explained below. Solid, filled symbols indicate the vowel offset, i.e. 100% duration. Finally, monophthongal vowel positions are also included for comparison of relative positions. Note that the diphthong plots in Figure 4.16 and below are scaled to cover a narrower region of F2 space than the vowel position plots in §4.1, so as to increase the resolution of the diphthong trajectories; this has the effect of excluding the vowel /i/, whose F2 value far exceeds all other vowels.

In general terms, every diphthong trajectory in Figure 4.16 begins in motion towards a low (for /aj, aw/) or back (for /ɔj/) target which is reached at or very near to 20% of its duration, indicated by the first open/unshaded symbol on each trajectory's plot line. In nearly every case the trajectory changes direction sharply at this point; the one obvious exception to this is pre-voiceless /aj/ which has a smaller degree of change occurring somewhat earlier than 20%. Towards vowel offset each trajectory changes in a central direction, typically fairly suddenly, at or near 90% of duration, indicated by the last open/unshaded symbol.

As will be seen, 20% and 90% of duration (or positions very close to these values) appear to mark important articulatory transitions fairly consistently across all three diphthongs, and all three datasets. Taking the 20% point of the trajectory to indicate the position of the first articulatory target, i.e. the nucleus, and 90% the second target of the diphthong, i.e. the glide or secondary vowel, each allophone can be transcribed phonetically using the nearest monophthongal vowel positions for the dialect in question. I attempt to provide transcriptions for each dialect which are fairly broad in terms of proximity to monophthongal positions, but narrow enough to highlight where pairs of allophones differ substantially from each other.

Winnipeg's diphthongs:

- /aj/: [aɪ] pre-voiced, [ʌɪ] pre-voiceless
- /aw/: [æʊ] pre-voiced, [ʌʊ] pre-voiceless
- /ɔj/: [oɪ] in both contexts

The overall findings from the Winnipeg dataset are largely compatible with Hagiwara's (2006) Winnipeg findings, as shown in Figure 4.15, but there are several points of contrast which can be noted; I cannot comment on the source of these differences, but can only note their existence. It is important to recall, too, that Hagiwara's trajectories only cover between 25%-75% duration, a span which is entirely internal to that covered between the 20% and 90% timepoints shown in Figure 4.16. Beginning with /aj/, Hagiwara's pre-voiced /aj/ appears to have a much lower and more advanced nucleus, quite close to but lower than /æ/ (the lowest vowel in the system); the present dataset does not indicate such a low or advanced nucleus. Hagiwara's pre-voiceless /aj/ appears to advance (at 75% duration) nearly as far forward as /i, e/; in the present dataset, its position is somewhat further back, not advancing forward past the position of /ɪ/ (although rising higher than /ɪ/).

For /aw/, Hagiwara's initial 25% positions are somewhat different for both allophones as compared to the results in Figure 4.16. Pre-voiced /aw/ in the present dataset starts much lower and more forward, very close to the position of /æ/, and the nucleus of pre-voiceless /aw/ is exactly in the position of /ʌ/. Additionally, in Hagiwara's data, towards vowel offset both allophones of /aw/ retract back further than the positions of /u, ɑ/, whereas in this study they retract towards a target position only as far back as /u/ at 90%, before turning in a central direction. A final point of contrast here concerns the momentum of pre-voiced /aw/. In Hagiwara's data, by the 50% timepoint its trajectory has covered a fair amount of distance, rising above the height of /ʌ/. In the present study, at 50% of duration pre-voiced /aw/ has covered much less distance and is still very low, below the heights of both /ʌ, ɑ/.

The trajectories of /ɔj/ are not drastically different between the two studies. The one point of contrast is that pre-voiceless /ɔj/ in the present dataset retracts further back for its nucleus such that the nuclei of both the pre-voiced and pre-voiceless allophones are fairly close to each other.

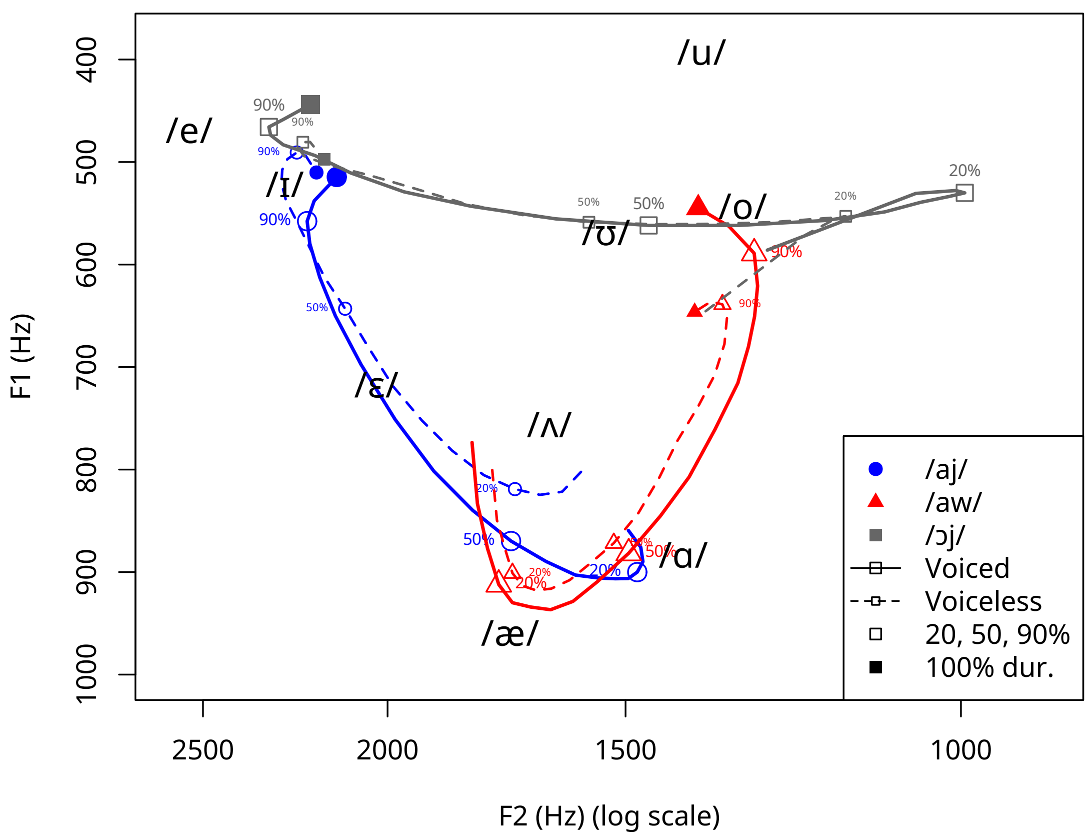


Figure 4.17 Diphthong trajectories: Denver

The diphthong trajectories from the Denver dataset are presented in Figure 4.17. The following phonetic transcriptions can be assigned based upon the 20% and 90% positions for each diphthong.

Denver's diphthongs:

- /aj/: [ɑɪ] pre-voiced, [ʌɪ] pre-voiceless
- /aw/: [æo] pre-voiced, [æo̹] pre-voiceless
- /ɔj/: [o̠ɪ] pre-voiced, [oɪ] pre-voiceless

The majority of Denver's diphthong trajectories are quite similar to those of Winnipeg. The nuclei of both allophones of /aj/ occur in similar acoustic positions as they do Winnipeg, but because of the general fronting of Denver's back vowels (see §4.1), the nucleus of pre-voiced /aj/ is very close to the position of Denver's /ɑ/. Pre-voiced /aj/ in Denver is more advanced than in Winnipeg, such that both allophones of /aj/ in Denver have offsets which are very close together; in Winnipeg, pre-voiced /aj/ is more retracted at 90% and thereafter, and pre-voiceless /aj/ advances (i.e. is fronted). The trajectories of both allophones of /ɔj/ are very similar in both cities, although Denver's are very slightly lower overall, and the nucleus of pre-voiced /ɔj/ in Denver extends relatively further back in comparison to the pre-voiceless nucleus than is the case for Winnipeg.

The one major difference between Denver and Winnipeg concerns /aw/, specifically its pre-voiceless allophone. In Denver, the two allophones of /aw/ have near-identical nuclei and follow very similar trajectories, with pre-voiceless /aw/ somewhat centralized along its entire trajectory and having a final position (both at 90% of duration and final offset) somewhat lower than pre-voiced /aw/. This relationship is much different in Winnipeg. In a very real sense, Winnipeg's pre-voiceless /aw/ is "raised" compared to pre-voiced /aw/, but this is not simply due to the nucleus itself being in a higher position (which it is); the entire trajectory of pre-voiceless /aw/ from onset until 90% of duration is shifted sharply upwards, only aligning with the pre-voiced allophone at offset. In summary, in Denver the trajectories of both allophones of /aw/ align very closely, with pre-voiceless /aw/ being centralized, until just before 90% of duration; the two allophones appear to have identical nuclei but different off-glide positions. In Winnipeg, the situation is reversed, as both allophones of /aw/ have different nuclei and different (if parallel) trajectories up to 90% of duration, with pre-voiceless /aw/ being substantially raised relative to pre-voiced /aw/; but, the two allophones have virtually identical offset positions for their off-glides.

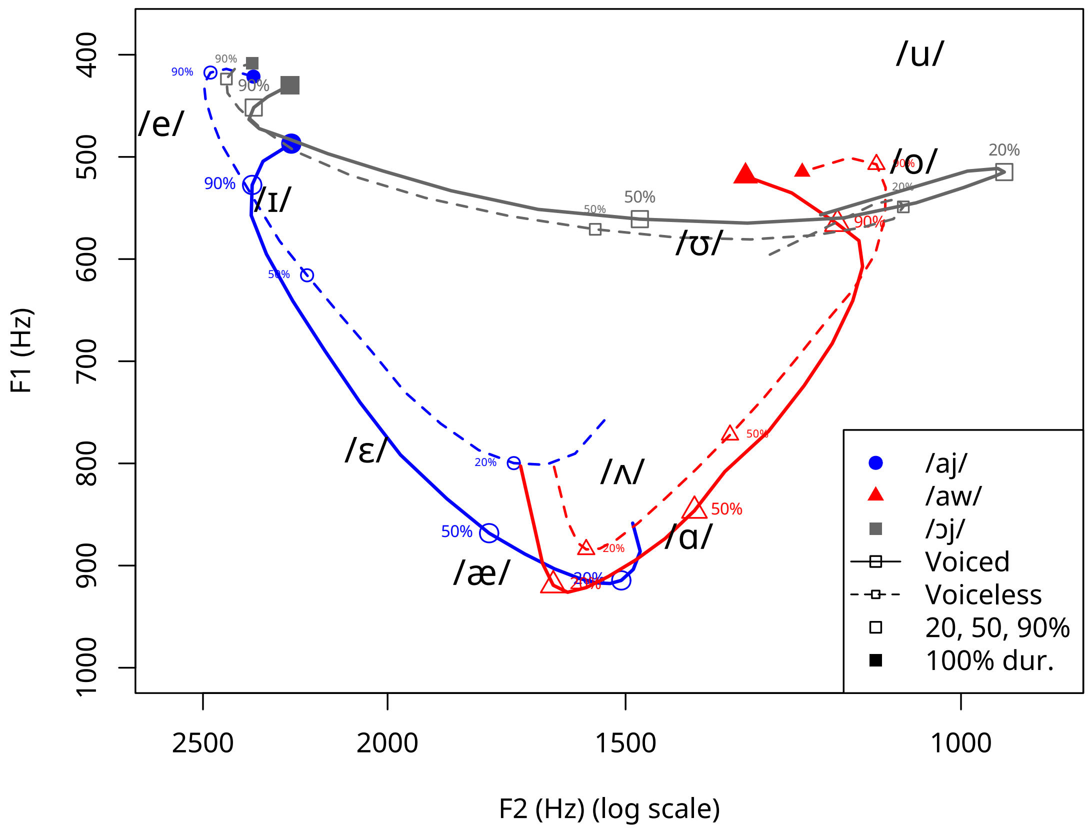


Figure 4.18 Diphthong trajectories: Madison

The diphthong trajectories from the Madison dataset are presented in Figure 4.18. The following phonetic transcriptions can be assigned based upon the 20% and 90% positions for each diphthong.

<u>Madison's diphthongs</u>:

- /aj/: [aɪ] pre-voiced, [ʌɪ] or [ʌe] pre-voiceless
- /aw/: [ao] in both contexts (possibly [aö] pre-voiced)
- /ɔj/: [oɪ] or [oe] in both contexts (possibly [o̱ɪ] or [o̱e] pre-voiced)

Madison's /aj/ allophones are at least as distinct from each other as are Winnipeg's, but the general pattern is the same as that in both of the other cities. The nucleus of pre-voiced /aj/ is low, between /æ/ and /ɑ/, that of pre-voiceless /aj/ is somewhat in advance of /ʌ/. Both have offset points well above the position of /ɪ/, with pre-voiceless intermediate between /ɪ/ and /e/. While there are somewhat idiosyncratic characteristics to the production of /aj/ in each city, it seems reasonable to characterize all three as exhibiting a version of pre-voiceless *raising* of /aj/.

The trajectories of the allophones of /ɔj/ in Madison are again remarkably similar to the other two cities, although pre-voiceless /ɔj/ is slightly lower across most of its trajectory, and the pre-voiced nucleus is retracted further than the pre-voiceless nucleus to about the same degree as in Denver, while Winnipeg exhibits more correspondence between the two nuclei in this respect. Overall, the production of /ɔj/ is highly similar across all three cities.

/aw/ in Madison occupies a position somewhat intermediate between Denver and Winnipeg. Like Denver, the two trajectories are fairly close and parallel, at least for the initial portion of the trajectory, although the pre-voiceless allophone is centralized to a somewhat greater degree in Madison. The off-glide and vowel offset, however, are more reminiscent of Winnipeg; at 90%, pre-voiceless /aw/ is raised higher than pre-voiced, and by offset the two occur very near to each other comparable with Winnipeg where they are in near-identical positions, and contrasting with Denver where the pre-voiceless glide and offset are much lower than the pre-voiced allophone. It might be said that in Denver /aw/ is not raised in pre-voiceless position, in Winnipeg it is raised, and in Madison it is "partially" raised; that is, pre-voiceless raising may not be as categorical as it is typically described.

4.4 COMPARING DURATIONALLY-DISTINCT FORMANT TRAJECTORIES

In §4.2 the occurrence of PVVA was established among the three dialects of English under investigation. In each dialect, ANOVA tests of vowel duration differences in relation to syllable type found the largest effect sizes among the three diphthongs /aj, aw, ɔj/ (with the exception of /aw/ in Madison, which is exceeded by /o/ in this respect; see Table 4.6 through Table 4.8). PVVA was further evaluated on the basis of the ratio of duration in pre-voiceless to pre-voiced coda contexts; /aw, ɔj/ had the largest ratios (smallest degree of PVVA) of the entire vowel inventory in each city, whereas the duration ratio of /aj/ was fairly close to the mean PVVA ratio among the vowel inventory of each city (see Figure 4.11 through Figure 4.13). In §4.3 the positions and trajectories of each of the three English diphthongs were examined in detail. In general, each diphthong exhibited distinct trajectories in pre-voiced vs. pre-voiceless coda contexts, often with different nuclei and/or off-glide positions. /ɔj/ typically exhibited the least amount of difference between its coda-context trajectories. Overall, the allophonic trajectories of both /aj, ɔj/ were remarkably similar across cities, while /aw/ exhibited the most substantial difference between cities in terms of its various allophonic realizations.

Taken together, these findings establish that diphthong duration is significantly correlated with coda voicing, that this effect varies in magnitude between diphthongs, and that acoustic patterns of diphthong production differ between dialects in ways that are not straightforwardly derived from these two facts. For example, although the duration ratios of /aw, ɔj/ are similar to each other and distinct from /aj/ within each dialect, cross-dialectally /aj, ɔj/ appear most similar in terms of the relationships of their allophonic trajectories, while /aw/ is more distinctive between dialects. In order to further examine the relationship between the coda-context allophones of the diphthongs both within and across the three dialects, a method of comparing durationally-distinct formant trajectories is required. This involves the application of time-scaling to the timepoint-based formant measurements gathered in each city, the implementation of two statistical techniques for comparing curvilinear data, and evaluation of cross-method results from those techniques; each of these steps are described individually in the following four subsections.

4.4.1 FORMANT TRAJECTORY TIME-SCALING METHODS AND PVVA MODELS

Recall from §3.3 that formant values were collected at 5% intervals across each vowel token's duration, yielding 20 timepoint-based measurements per formant. In §4.3 these timepoint-based formant values were plotted as intersections in F1×F2 space—a proxy for articulatory position—mapping the trajectories correlated with the articulatory path of each coda-context allophone. This method of visualization, while portraying durational differences to some degree via the relative lengths of each trajectory, does not provide a principled means of comparing trajectories with different durations. As PVVA has been established to reliably occur for all vowels (with the sole exception of /æ/ in Madison; see §4.2), it seems essential to develop a method which incorporates durational information more directly, along with comparison of formant values, as a proxy for articulatory position.

The method used to compare formant trajectories in §4.3 involves plotting the intersection of F1 and F2 values across the duration of each allophone. Detailed resolution of differences in acoustic (and by implication, articulatory) position may be achieved by highlighting various points along the trajectory which are visually notable; this allowed identification of e.g. the 20% point as an important position potentially representing the diphthongal nucleus. However, this approach is somewhat ad hoc in nature, and less overtly obvious differences may go undiscovered via this methodology. An alternative to this method involves segregation of the F1 and F2 trajectories of each allophone. Although articulatory position is less easily interpreted when formant values are not combined, which is the main advantage of the method used in §4.3, this segregation permits a more principled means of comparison across different-duration allophones via statistical methods which are discussed in subsequent subsections below. As will be seen, there are at least two distinct ways to carry out such a comparison of segregated formant trajectories, each having specific implications for how PVVA is achieved or carried out.

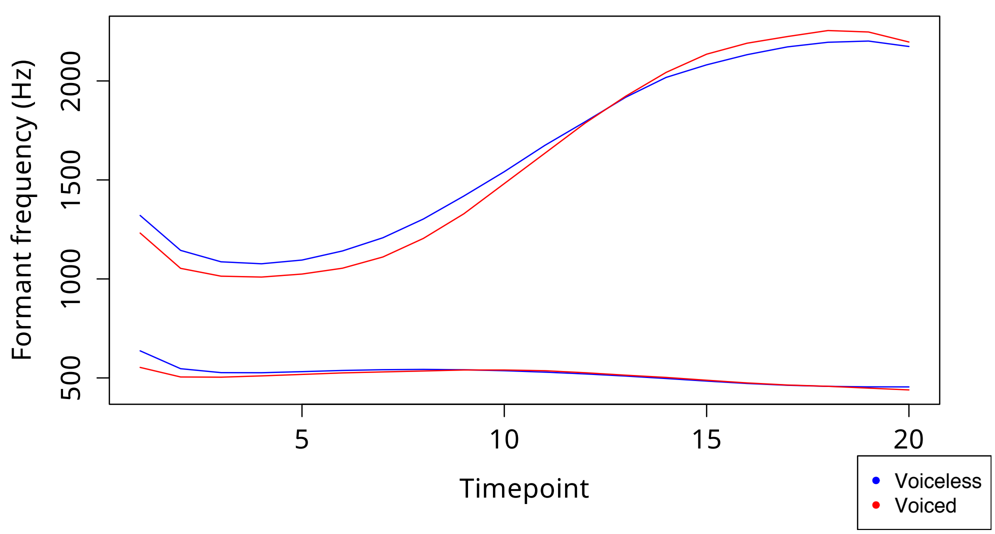


Figure 4.19 Formant trajectories of /ɔj/ by coda voice, time-normalized duration-scaling: Winnipeg

The first method of formant comparison involves the plotting of (segregated) formant trajectories in *time-normalized* fashion, matching each timepoint (t1, t2, t3 … t20) directly between the two sets of data, and ignoring the actual durational differences between the two allophones. As an example, Figure 4.19 illustrates a combined plot of the mean values of F1 and F2 for both the pre-voiced and pre-voiceless allophones of /ɔj/ (from the Winnipeg dataset) over their full duration. The close correspondence between the pre-voiced (red) and pre-voiceless (blue) formant trajectories indicates that the acoustic (and thus articulatory) paths of both allophones of /ɔj/ are quite similar despite being durationally distinct; this coincides with the previous findings for /ɔj/ in Winnipeg, which indicated that the two allophones have similar articulatory trajectories (see §4.2 and §4.3).

This correspondence under the *time-normalization* method indicates that the achievement of PVVA for /ɔj/ can be described under a model of either articulatory *compression* or *expansion*. Under each of these models, the same articulation occurs before both voiced codas and voiceless codas, but at differing rates in each case. The choice of articulatory *compression* vs. *expansion* depends on the selection of the assumed default form: the pre-voiced or pre-voiceless allophone. The *compression* model assumes that the *pre-voiced* allophone is the default form, whose articulation is *compressed* before a voiceless coda, occurring over a briefer time period while still covering the same articulatory path. The *expansion* model assumes that the *pre-voiceless* allophone is the default, whose articulation is *expanded* before a voiced coda, occurring over a lengthier time period while still covering the same articulatory path. In either case, the observed correspondence between the two sets of formant trajectories remains the same. As such, determination of which allophone should be assigned default status, and by implication which model, compression or expansion, is most appropriate, cannot be made on the basis of trajectories which match under a time-normalization comparison, such as /ɔj/.

In addition to compression/expansion, another available model for achieving PVVA is articulatory *truncation*, wherein a portion of the default articulation is eliminated in the production of the abbreviated allophone. Unlike the compression vs. expansion models discussed above, the *truncation* model is not ambiguous with respect to the choice of default allophone, as truncation involves removal of some part of the default articulation, implying that the resulting form will be durationally abbreviated rather than lengthened. Thus, the durationally longer form, i.e. the *pre-voiced* allophone, must be considered the default in the case of the truncation model. As truncation may apply to either the initial or final portion of an articulation, there are two sub-models here: *truncation of onset*, where the initial portion of the articulation is truncated and the medial and offset portions are similar across both allophones; and, *truncation of offset*, where the final portion of the articulation is truncated and the onset and medial portions are similar.

A method of time-scaled formant trajectory comparison which is compatible with the truncation model (and sub-models) involves *proportional-scaling* rather than *time-normalization*. Under this method, the durations of the formant trajectories of pre-voiceless allophones are scaled down according to their overall proportion of their duration in relation to the duration of pre-voiced allophones. In terms of implementation, this was achieved by restricting the comparison to include all of the pre-voiceless allophone's formant data, but for the pre-voiced allophone only a duration-matched portion of the data was retained. If the pre-voiceless allophone's mean duration was 60% of that of the pre-voiced allophone, then 60% of the pre-voiced formant data points (either the first or the last 60%) were included, and the remainder excluded. The included timepoints were assigned new, scaled durational values such that their full durational span matched that of the entirety of the pre-voiceless allophone's formant data points.

For example, if a pre-voiceless allophone had a mean duration of 600 ms and its pre-voiced allophone had a mean duration of 1000 ms, then only a 600 ms portion of the pre-voiced allophone's data points were included. As 20 measurement points were included per formant per token, the pre-voiced allophone would have measurements at every 1/20th, or 5% of 1000 ms: at 50, 100, 150, 200 … to 1000 ms. The pre-voiceless allophone, on the other hand, would have measurements at every 5% of 600 ms: at 30, 60, 90, 120 … 600 ms. To compare these under *proportional-scaling*, only a 600 ms portion of the data points of the pre-voiced allophone would be included. To compare the *initial* 600 ms portion of the pre-voiced allophone, the data points from 50 (initial measurement) to 600 ms would be retained and compared to the full set of data points from the pre-voiceless allophone, while the data points subsequent to 600 ms would simply be excluded. To compare the *final* portion of the pre-voiced allophone, however, the last 600 ms portion would be retained, spanning durational values from 400 to 1000 ms. As the pre-voiceless data points would have values only as high as 600 ms, in order to implement such a comparison, the pre-voiced durational values were recalculated by subtracting the difference between the durations of the two allophones, in this case 1000 – 600 = 400 ms, so that e.g. a data point at 650 ms would be reassigned a durational value of 650 – 400 = 250 ms. These recalculated durations, as illustrated in Table 4.9, would then be used to compare the two sets of formant data points.

Table 4.9 Comparing proportionally-scaled formant data, final portion of pre-voiced allophone compared to entirety of pre-voiceless allophone

Timepoint	Pre-voiceless	Pre-voiced; original values	Pre-voiced; scaled values
1	**30 ms**	50 ms	*Excluded*
2	**60 ms**	100 ms	*Excluded*
3	**90 ms**	150 ms	*Excluded*
4	**120 ms**	200 ms	*Excluded*
5	**150 ms**	250 ms	*Excluded*
6	**180 ms**	300 ms	*Excluded*
7	**210 ms**	350 ms	*Excluded*
8	**240 ms**	400 ms	400 – 400 = **0 ms**
9	**270 ms**	450 ms	450 – 400 = **50 ms**
10	**300 ms**	500 ms	500 – 400 = **100 m**s
11	**330 ms**	550 ms	550 – 400 = **150 ms**
12	**360 ms**	600 ms	600 – 400 = **200 ms**
13	**390 ms**	650 ms	650 – 400 = **250 ms**
14	**420 ms**	700 ms	700 – 400 = **300 ms**
15	**450 ms**	750 ms	750 – 400 = **350 ms**
16	**480 ms**	800 ms	800 – 400 = **400 ms**
17	**510 ms**	850 ms	850 – 400 = **450 ms**
18	**540 ms**	900 ms	900 – 400 = **500 ms**
19	**570 ms**	950 ms	950 – 400 = **550 ms**
20	**600 ms**	1000 ms	1000 – 400 = **600 ms**

Duration values used for proportional-scaling comparison in boldface

This methodology is perhaps most easily illustrated via a concrete example. In Figure 4.20, the formant trajectories of pre-voiceless /aj/ (Winnipeg) are *proportionally-scaled* according to their duration in relation to the duration of pre-voiced /aj/, a ratio of 0.577 (see Table 4.6). In this case, the *truncation of onset* model is most applicable, so in Figure 4.20 the scaled-down pre-voiceless trajectories are *aligned at the right edge* as indicated by the blue arrow, as if the initial (left) portion of the pre-voiced articulation were truncated, and the articulation of each allophone coordinated towards the vowel offset position. The numerical timepoints at the bottom of the figure refer to the timepoints of the pre-voiced allophone.

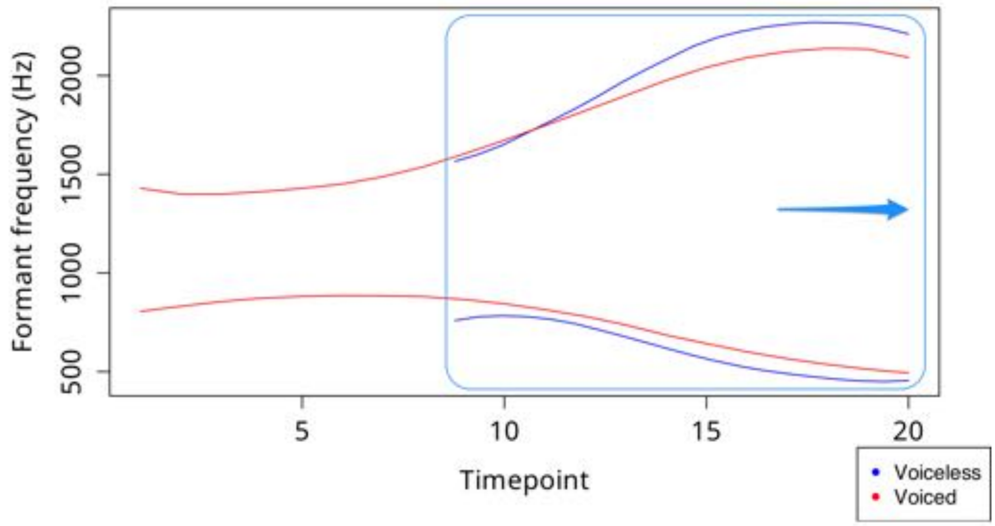


Figure 4.20 Formant trajectories of /aj/ by coda voice, proportional duration-scaling, right-alignment: Winnipeg

Although the correspondence between the formant trajectories in Figure 4.20 is somewhat less cohesive than that observed for /ɔj/ in Figure 4.19, it is much superior to the alternative methods of time-scaling, *time-normalization* and *proportional-scaling with left-alignment*.

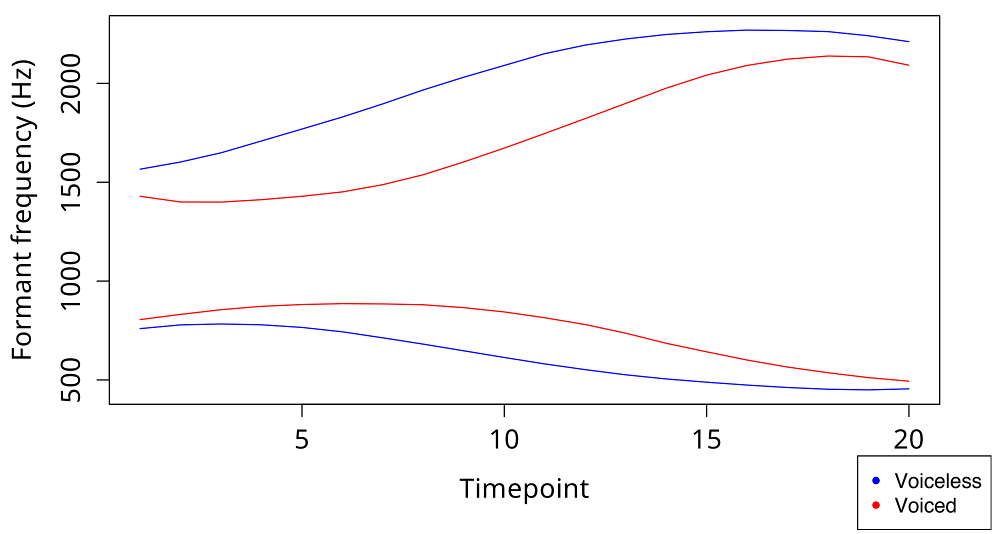


Figure 4.21 Formant trajectories of /aj/ by coda voice, time-normalized duration-scaling: Winnipeg

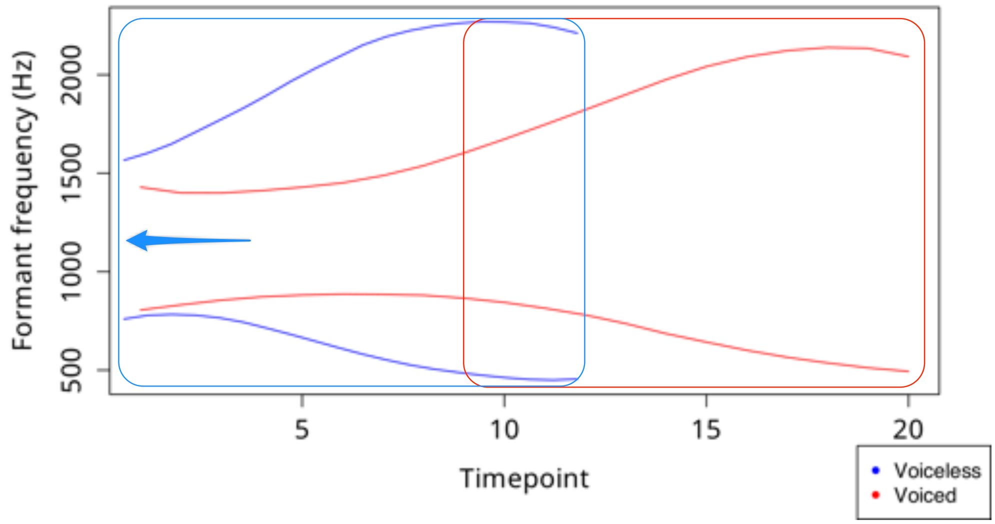


Figure 4.22 Formant trajectories of /aj/ by coda voice, proportional duration-scaling, left-alignment Winnipeg

A *time-normalization* comparison, such as was compatible with /ɔj/, is shown for /aj/ in Figure 4.21. And, *proportional-scaling with left-alignment* or coordination at vowel onset, i.e. *truncation of offset*, is shown in Figure 4.22. As should be clear from these examples, *proportional-scaling with right-alignment*, or *truncation of onset* as shown in Figure 4.20 is by far the superior method in the case of Winnipeg /aj/. As noted above, articulatory *truncation* is only available as a means of achieving PVVA if pre-voiced allophones are taken to be the default form, as they are durationally longer and contain articulatory events which are absent in their pre-voiceless counterparts, and thus available for truncation. The converse is not true; derivation of the longer, pre-voiced allophones from the pre-voiceless allophones would require another distinct mechanism such as the *addition* of some other articulatory event not present in the default articulation. This contrasts with *compression* vs. *expansion* models, neither of which fundamentally alters the articulatory content of either allophone.

On the whole, articulatory *truncation* appears to be a simpler process than *addition*, involving only the removal of existing articulatory components vs. the incorporation of novel components. Taking the pre-voiced allophone to be the default form is also consistent with traditional interpretations e.g. of Canadian Raising, where raising only occurs before voiceless codas, and where voiced codas and open syllables are understood to be the "elsewhere" environment and thus indicative of the default form. Additionally, as seen in §4.2 pre-voiced vowels and vowels in open syllables are more similar to each other durationally than either is to pre-voiceless vowels, nearly across the board (/æ/ being the sole exception). Durationally-abbreviated forms occur in a more restricted environment, and are thus also unlikely to represent the default. For these reasons, I assume hereafter that PVVA is best described as a process which *abbreviates* vowels in pre-voiceless position rather than *lengthening* them in pre-voiced position. I also hereafter refer to the three comparative methods of formant trajectory time-scaling, *time-normalization*, *proportion-scaling with left-alignment* and *proportional-scaling with right-alignment* as time-scaling *models*, implying their associated PVVA models: *compression, truncation of offset*, and *truncation of onset*, respectively. Vowels such as Winnipeg /ɔj/, whose pre-voiced and pre-voiceless articulations correspond well under a *time-normalization* comparison may thus be interpreted as achieving PVVA via *compression*. Likewise, vowels such as Winnipeg /aj/, whose pre-voiced and pre-voiceless articulations correspond well under a *proportionally-scaled with right-alignment* comparison may be interpreted as achieving PVVA via *truncation of onset*. The different time-scaling models favoured by Winnipeg's /ɔj/ and /aj/ suggest that PVVA may be achieved by multiple articulatory means even within the same dialect.

The comparisons presented thus far are only rudimentary, however. In order to fully describe the ways in which PVVA occurs throughout the entire vowel systems of multiple dialects, the data must be subject to more complex statistical modelling than the simple comparison plots of formant trajectories presented above. Two contemporary tools used in the analysis of curvilinear data (e.g. formant trajectories) are the *smoothing spline analysis of variance* (SSANOVA: Gu & Wahba 1993; Gu 2002; Davidson 2006; Baker 2006; Chen & Lin 2011) and *generalized additive mixed models* (GAMMs: Hastie & Tibshirani 1987, 1990; Buja, Hastie & Tibshirani 1989; Wood 2004, 2006; Sóskuthy 2017). These types of tools allow statistical comparison between curves—SSANOVA “splines” or GAMMs “smooths”—derived from multiple datasets. Comparison between paired splines/smooths permits the inference of significant differences or similarities between the source datasets, e.g. between vowel formants in voiced vs. voiceless coda contexts. Due to the innovative nature of the particular application of these statistical tools used in this study, as is explained below, I felt that it was desirable to determine the results via multiple methods, hence the use of both SSANOVA and GAMMs, and then cross-tabulate the results. In the following two subsections, the application of each of these tools is described, with the cross-evaluation following in the final subsection.

4.4.2 SSANOVA COMPARISONS OF FORMANT TRAJECTORIES

Of the two statistical methods discussed in §4.4.1 for handling comparisons of nonlinear data, SSANOVA is probably the one most generally familiar to linguists. This technique was first brought to the attention of the wider linguistics community by Davidson (2006), who applied it to the analysis of ultrasound measurement of tongue shapes. That particular application involved examination of static, non-dynamic but complex curvilinear shapes (of the tongue). As discussed by Davidson, SSANOVA is applicable to any nonlinear dataset including those which represent measurements of single points, e.g. formant values, that change position over time, as it can be equally used to "investigate comparisons along spatial and *temporal* dimensions," (Davidson 2006:409, emphasis added).

One of the appeals of SSANOVA is its ease of interpretability. The *smoothing splines* produced by SSANOVA are mathematical curves, derived from and representative of a particular dataset. Modern SSANOVA implementation augments smoothing splines with the addition of Bayesian confidence intervals, which allow visual comparisons to be conducted between splines representing different datasets: "when the intervals cover about 95% of the values of the true curve at the data points, the intervals more or less 'graze' the truth, and the width of the intervals is visually interpretable by an unsophisticated user as an accuracy indicator," (Gu & Wahba, 1993:99). When confidence intervals of two distinct splines do not overlap, this can be taken as evidence of a significant ($p<0.05$ for 95% intervals) difference between the two represented datasets. Conversely, when the splines do overlap there is no evidence for a significant difference between them, and hence between the two datasets. It's important to remember, however, that lack of statistical evidence for a difference in a given case does not definitively indicate that such a difference does not exist, but merely that it cannot be demonstrated by the applied test or measure, and that it may potentially be revealed by other means.

Beyond its use for ultrasound measurements and other types of static measurements, SSANOVA techniques have more recently been implemented in the analysis of formant trajectory data. A particularly relevant example here is Hall (2016a,b) who applied the technique to data on the Canadian Raising diphthongs /aj, aw/ as produced by speakers in the Canadian cities of Toronto and Vancouver.

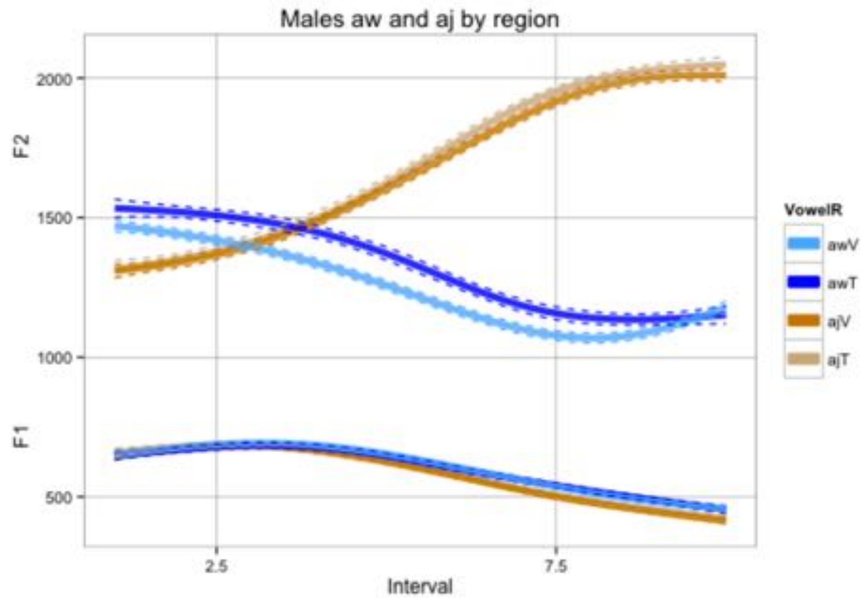


Figure 4.23 SS-ANOVA results for males, /aw/ and /aj/ by region (Hall 2016b:29)

Figure 4.23 illustrates the differences between the formant trajectories of the two diphthongs for Hall's (2016b) male Vancouver ('V') and Toronto ('T') speakers. Smoothing splines are indicated as solid coloured lines, and confidence intervals as dashed lines of the same colour; here, these form very tight boundaries adjacent to the splines, in some cases, e.g. F1 of /aj/, so close that they virtually overlap the visible splines themselves. Focusing on the F2 (upper) splines for /aw/ (blue), it can be observed that the confidence intervals for Vancouver (pale blue) vs. Toronto (dark blue) male speakers do not overlap for the majority of the trajectory, with bare near-overlap at vowel onset, and a brief period of complete overlap at offset. This indicates that, for the bulk of the vowel's durations, /aw/ has different F2 values for speakers in the two cities, indicating different degrees of articulatory frontness of /aw/ over time between the two populations. In contrast, the F2 values for /aj/ (reddish-orange splines) are overlapped or nearly overlapped between the two sets of speakers across virtually the entire duration, indicating that there is no significant F2 difference between the two groups across the majority of the trajectory, and therefore implying no significant difference in degree of frontness for the majority of the articulation of /aj/ over time. As this example illustrates, SSANOVA is a useful technique for the comparison of temporal, dynamic and non-linear features, such as acoustic formant trajectories.

Implementation of SSANOVA with the Winnipeg, Denver, and Madison datasets proceeded as follows. Formant measurements were compared between voiced and voiceless coda contexts for each vowel. Recall from §4.4 that three models of time-scaling were identified for comparing durationally distinct allophones in relation to coda voicing: *time-normalized* vs. *proportionally-scaled*, with *left-alignment* and *right-alignment* variations for the latter. SSANOVA comparisons were run in R for each of the three models, for F1 and F2 among all twelve vowels (excluding /ʊ/) in the three datasets. The R script used for this function was based on Wassink (2013), with substantial modification, and relies on the R package `gss` (Gu 2014) for implementation.

As noted earlier, the general function of an SSANOVA is to determine whether or not there is a significant difference between two groups of data, as represented by the smoothing splines and their associated confidence intervals, which is done by inspecting for regions of overlap. Even when making a single comparison, such as comparing one formant of one vowel across two conditions, this task is not completely straightforward. For example, as noted earlier, in Figure 4.23 the splines and confidence intervals for F2 of /aj/ are nearly completely overlapped between Vancouver and Toronto speakers—but not entirely. There is some slight separation of the two just past the 50% mark, and near separation towards vowel offset. Determining the extent to which such a difference matters is not a simple task.

In the case of the data under examination here, the overall complexity of the task of SSANOVA interpretation is a function of the inclusion of multiple formants (both F1 and F2) in multiple vowels (twelve) across three cities and three different alignment models. The implementation of SSANOVA here is also atypical and innovative. In more typical applications such as Hall (2016a,b), SSANOVAs are used to determine the degree to which some feature, e.g. a particular vowel formant, differs across multiple conditions, e.g. differently-voiced codas. Here instead, SSANOVA testing is used to determine which model, on a per-vowel basis, is the best fit for the data in that it *minimizes* the difference across the two conditions, rather than strongly indicating a statistical difference between them; the assumption behind this being that the two durationally distinct allophones of each vowel are related and derive from some common (articulatory) origin. Although one of the two allophones can be taken as the default or underlying form (and I have argued in §4.4.1 that it seems reasonable to take this to be the pre-voiced allophone), it isn't actually necessary to specify the default in order to carry out statistical comparisons to identify commonalities between the two forms. The SSANOVA analyses conducted here (and the GAMMs analyses conducted in §4.4.3) do not in fact make any such specification, but merely compare formant trajectories across the two coda voicing conditions under the various time-scaling models. The ideal SSANOVA time-scaling model for a given pair of allophones, under this view, would produce splines and confidence intervals for each condition (i.e. coda voice quality) which overlapped maximally, relative to the other models being tested. An outcome wherein one model had completely overlapping splines (and confidence intervals), while the other two models had clearly distinct splines (and confidence intervals) would thus be the clearest case of the superiority of one model vs. the other two models.

The individual SSANOVA comparisons for each city's dataset are compiled in full in Appendices B through D. Here I will discuss some points of commonality and contrast, focusing on the diphthongs. There is remarkable agreement between the three dialects with regard to /aj, ɔj/ whose SSANOVA outputs are extremely similar and whose best-fitting time-scaling models are cross-dialectally identical: proportional-scaling with right-alignment for /aj/, and time-normalization for /ɔj/. The results in §4.3 also indicated that these two diphthongs had remarkably similar patterning with regard to their allophonic acoustic trajectories across dialects.

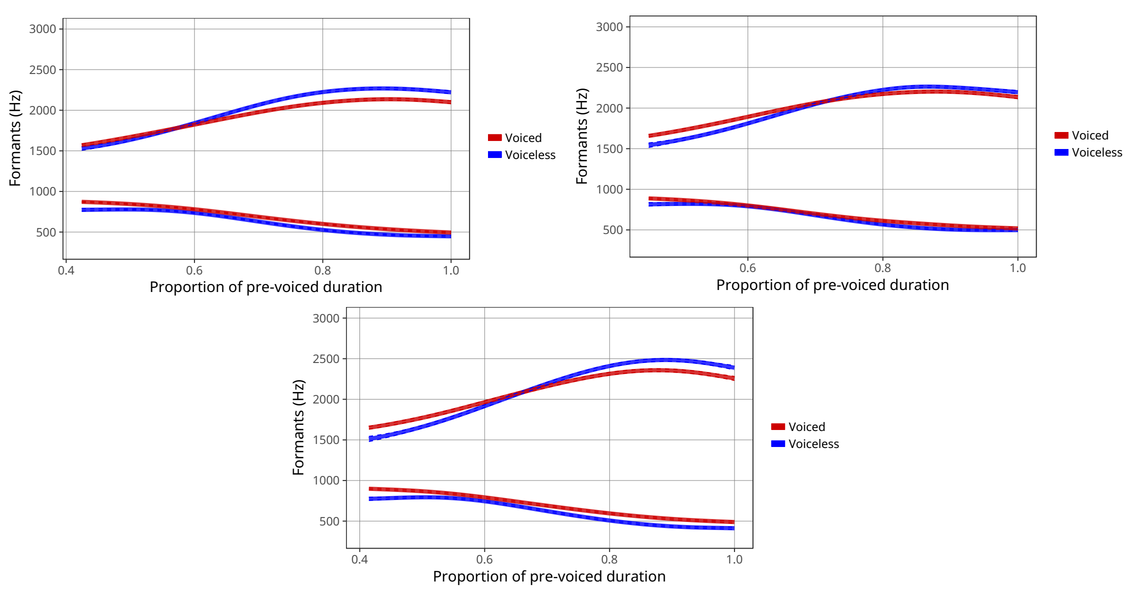


Figure 4.24 SSANOVA of /aj/ by coda voice, proportionally-scaled with right-alignment: Winnipeg (top left), Denver (top right), and Madison (bottom)

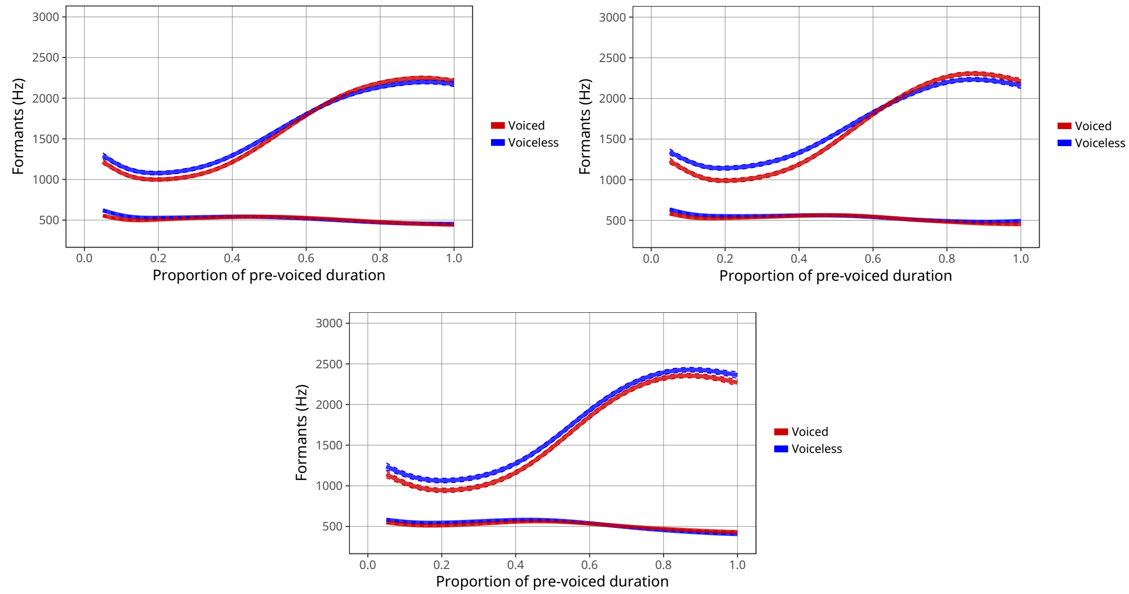


Figure 4.25 SSANOVA of /ɔj/ by coda voice, time-normalized: Winnipeg (top left), Denver (top right), and Madison (bottom)

Multi-dialect SSANOVA comparisons for /aj/ are displayed in Figure 4.24, and in Figure 4.25 for /ɔj/, illustrating the degree of correspondence between the three dialects for these two diphthongs. Note the distinction between proportional-scaled comparisons (Figure 4.24) and time-normalized (Figure 4.25); the former involves comparison of a duration-equivalent portion of the pre-voiced trajectory matched to the full-duration pre-voiceless trajectory, while the latter involves the full trajectories of both conditions.

The comparisons illustrated in Figure 4.24 and Figure 4.25 indicate strong congruence between all three dialects with respect to /aj/ and /ɔj/. In contrast to these, /aw/ exhibits a strong degree of correspondence between Denver and Madison, where the best-fitting model is time-normalization, but in Winnipeg proportional-scaling with right-alignment offers a somewhat superior fit.

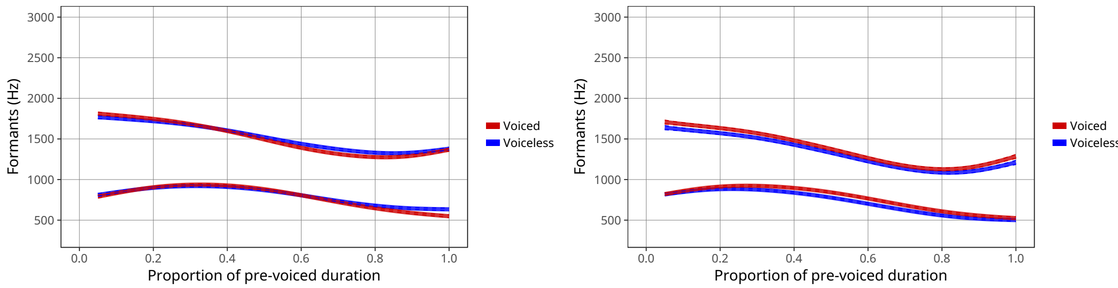


Figure 4.26 SSANOVA of /aw/ by coda voice, time-normalized: Denver (left) and Madison (right)

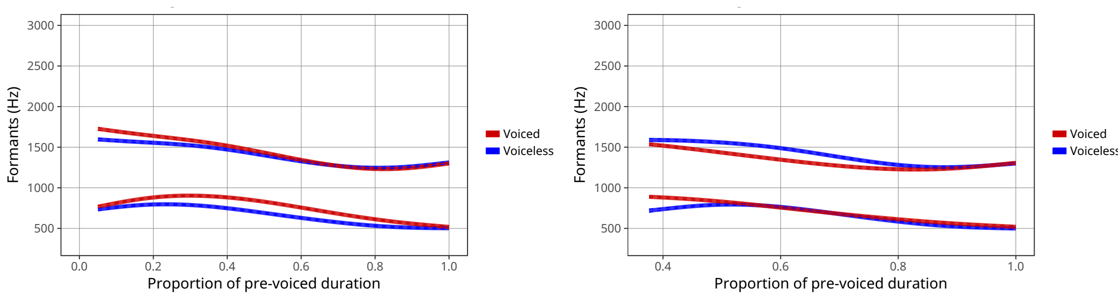


Figure 4.27 SSANOVA of /aw/ by coda voice in Winnipeg: time-normalized (left) and proportionally-scaled with right-alignment (right)

Figure 4.26 illustrates the SSANOVA comparisons with time-normalization for /aw/ in Denver and Madison, and Figure 4.27 compares the time-normalization and proportionally-scaled, right-aligned models[30] for /aw/ in Winnipeg. The distinction between the two SSANOVA models of Winnipeg /aw/ in Figure 4.27 is subtle; the main difference concerns the overall degree of spine overlap, which is slightly greater in the proportionally-scaled, right-aligned model. Each of the two models shown presents a sustained period of non-overlap—in F1 under time-normalization, in F2 under right-aligned proportional-scaling—and so neither presents as close of a fit as the results for

[30] The proportionally-scaled, left-aligned comparison is not illustrated here for Winnipeg as it is definitively inferior to either alternative; see Appendices B–D for full comparisons of all three models across the full vowel inventory of each dataset.

either Denver or Madison in Figure 4.26. Even if one were to take the cautious stance that a clear distinction cannot confidently be made between these two models for /aw/ in Winnipeg, this nonetheless presents a distinct situation from the two American dialects, where time-normalization is the clearly superior model; the case of /aw/ in Canada is at the very least a nuanced one, whereas in both American dialects it is clear-cut. This contrasts with the findings in §4.3, where the allophonic acoustic trajectories of /aw/ did not adhere to a strict cross-dialectal pattern, but rather Madison appeared to occupy somewhat of a middle ground between Winnipeg and Denver. Here the patterning (Winnipeg vs. Denver and Madison) is more robust, although it may be noted that the spline overlap for Denver is more substantial than Madison in Figure 4.26.

It is worth emphasizing that the application of SSANOVA via multiple models of time-scaling as described here is unorthodox, and I am unaware of this particular methodology, or a similar one, having been applied elsewhere within linguistic research. Therefore, GAMMs comparisons were also conducted in a similar fashion to the SSANOVA tests discussed above, in order to verify determinations of best-fitting time-scaling models across both statistical methods.

4.4.3 GAMMs COMPARISONS OF FORMANT TRAJECTORIES

GAMMs and their primary theoretical foundation GAMs (*generalized additive models*) have a decades-long history in statistical analysis, although within contemporary linguistics they may be less widely known and utilized than SSANOVA. The purpose behind the development of GAMs was to "provide a flexible method for identifying non-linear covariate effects in a variety of modeling situations," (Hastie and Tibshirani

1987:385), curvilinear formant trajectories being an example of such an application. The additional 'M' in GAMMs indicates the incorporation of *mixed modeling* within GAMs. Mixed modeling techniques combine fixed effects models (e.g. traditional linear regressions models) with random effects models, such that "some of the unknown coefficients (or functions) in the model linear predictor are now treated as random variable (or functions)" (Wood 2006:xii). Random effects and mixed models have risen in prominence within linguistics over the past decade or more, perhaps most notably in quantitative sociolinguistic research as well as other sub-fields (Johnson 2009; Tagliamonte & Baayen 2012; Gries 2015).

A discussion of the statistical underpinnings and merits of linear models, GAMs, GAMMs, random effects and mixed models is well beyond the scope of this dissertation. For GAMMs in particular, the interested reader is guided towards Sóskuthy (2017) which focuses on application to acoustic phonetics research. The GAMMs script (see Appendix D) used with the present datasets was compiled by myself, based largely on the methods described in Sóskuthy (2017), and relies on R packages `mgcv` (Wood 2011) and `itsadug` (van Rij, Wieling, Baayen & van Rijn 2016) for computation.

GAMMs appear to be less resource-intensive than SSANOVA methods, at least based upon my own experience in implementing the two procedures with the data from this dissertation. In practical terms, this means that more effects can be included in a given model, and/or run over larger datasets, as computation time and processing power needs are relatively reduced.

The SSANOVA models discussed in §4.4.2 were computed based on a single set of comparisons: the correlation between formant values, measured at each timepoint (time-normalized or proportionally-scaled), and coda voice quality. In GAMMs, “smooths”—roughly equivalent to SSANOVA splines—can be calculated for a variety of specified factors or effects. The GAMM function applied to the three datasets for Winnipeg, Denver, and Madison tested the correlation between the factors of *formant* value and *coda voicing*, with separate smooths calculated on the basis of *timepoint* generally, on the basis of *timepoint* with the data split according to *coda voice quality*, and with additional smooths calculated for *speaker* as a random effect (random smooth), allowing for individual variation between speakers to be incorporated into the model. The combination of the various smooths is then compared across the two coda voice contexts to determine, as with SSANOVA, whether they differ significantly.

GAMMs were calculated for each formant per vowel across the three datasets, run on duration-scaled data following the same three time-scaling models which were used in the SSANOVA testing in §4.4.2: time-normalized, proportionally-scaled with left-alignment, and proportionally-scaled with right-alignment. As with the SSANOVA models, the three GAMMs run per formant per vowel were compared in order to determine the best-fitting model in each case.

Unlike SSANOVA, GAMMs permits two methods of comparison. The first is similar to SSANOVA, and involves visual comparison of *difference smooths*, which are plots of the difference between the calculated GAMMs smooths for two conditions. As an example, Figure 4.28 illustrates the set of smooths calculated for F2 of /aj/ in Winnipeg.

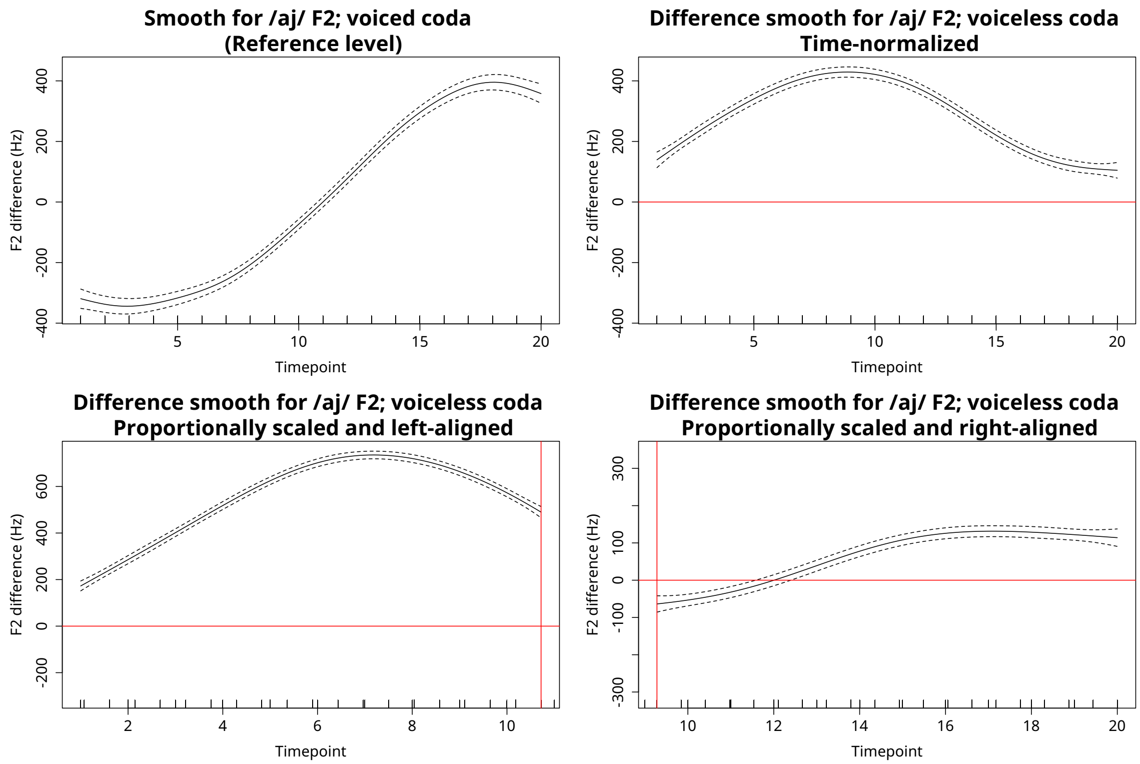


Figure 4.28 Difference smooths for F2 of /aj/: Winnipeg

The top left plot is the *reference level* smooth, which is set as the voiced coda condition[31], i.e. the longer duration allophone of /aj/. The other three plots are the *difference smooths* for the voiceless coda allophone under each of the three scaling/alignment models. Where and how far the difference smooth deviates from the zero-line in the plot indicates the degree to which the two formant trajectories are dissimilar, with the dashed-line confidence intervals surrounding the smooth line indicating the degree of certainty associated with its position. Both time-normalization (top right) and proportionally-scaled with left-alignment (bottom left) models exhibit large differences across their respective trajectories, with no

[31] For the purpose of running the GAMMs comparison here, the choice of which condition (i.e. coda voicing context) to use as the reference level is arbitrary; had the voiceless coda condition been selected as the reference level instead, the difference smooth plots would be mirror images of those shown in this section, inverted along the y-axis.

regions crossing the zero-line, i.e. the values of F2 across the two coda conditions do not overlap at all. The scaled, right-aligned plot (bottom right) indicates far less deviance overall, as well as having a region of zero-line overlap (i.e. maximal similarity with the reference level), indicating that this is the most suitable model in this case.

In addition to visual comparisons of difference smooths, different GAMMs models may also be compared against each other via a function which ranks them according to a value termed the Akaike Information Criterion (AIC):

> "AIC is a combination of two quantities: how surprising the data are given our fitted model (the lower this number, the better the fit) and how many parameters are used in the model. That is, AIC penalizes both bad model fits and unnecessary model complexity. When comparing two models, the one with a lower AIC should be preferred," (Sóskuthy 2017:26).

For F2 of /aj/ in Winnipeg, the right-alignment model was deemed superior to the time-normalization model by an AIC difference of –308.99, and superior to the left-aligned model by an AIC difference of –354.18, which are relatively large values that correspond well with the visual comparison illustrated in Figure 4.28.

Another method of visualization for GAMMs involves a combined plot of both smooths (plus confidence intervals) under comparison; such comparison plots end up visually resembling the typical SSANOVA format as seen in §4.4.2. An example of this type of plot is shown in Figure 4.29 F2 of /aj/ from the Winnipeg dataset, the same data used to produce the difference smooths in Figure 4.28.

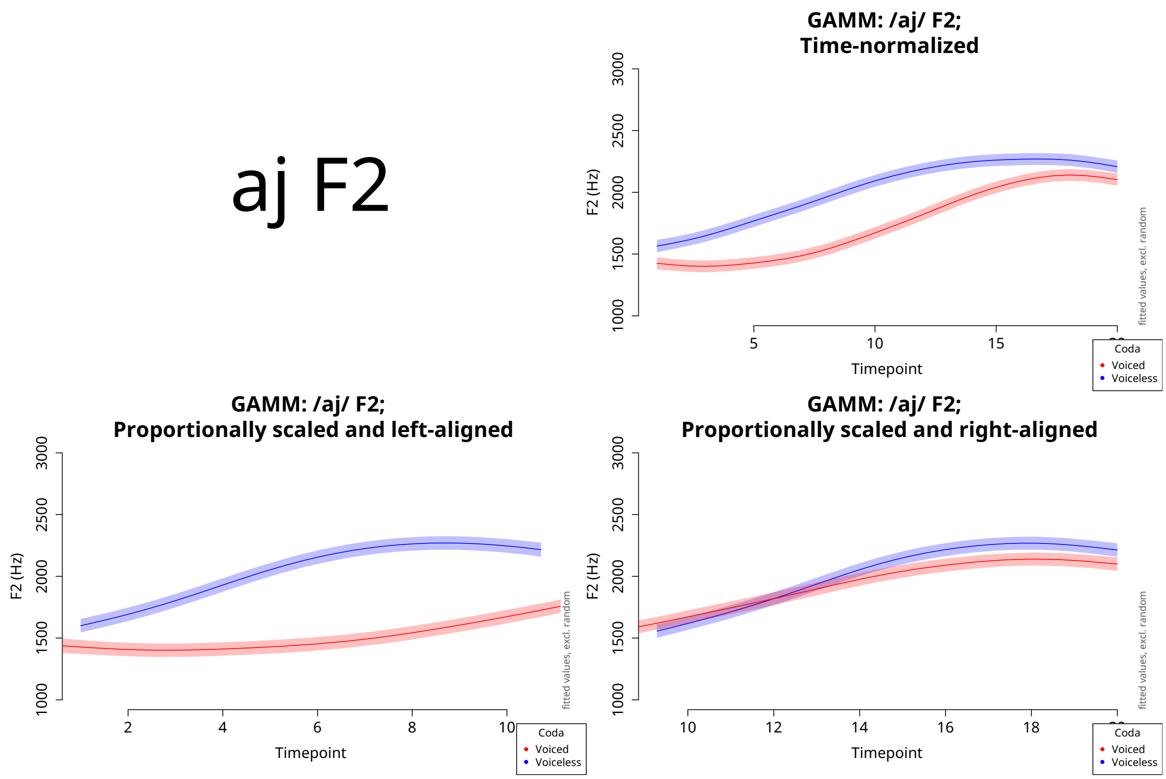


Figure 4.29 Smooths comparisons for F2 of /aj/: Winnipeg

While visually intuitive, Sóskuthy cautions against overreliance on this particular interpretation, however, noting two important points. First, the degree of *non*-overlap between confidence intervals required to determine significance is essentially arbitrary. While total non-overlap unequivocally indicates a significant difference between two smooths, partially overlapping cases are less clear. Second, and perhaps more importantly, the *presence* of overlapping confidence intervals across different smooths does *not* guarantee a non-significant difference between them: "we cannot make any conclusions about the significance of the difference when the confidence intervals do overlap: it may be significant but it may also be non-significant" (Sóskuthy 2017:17). Although difference smooths may be less visually intuitive than smooths comparisons, as they do not display formant values directly, according to Sóskuthy they do not suffer from either of the issues mentioned above. As such, although the combination smooths plots are included in the full set of GAMMs comparisons in Appendices F through G for illustrative purposes, only the difference smooths and AIC difference calculations were used in determination of best-fit GAMMs models.

4.4.4 EVALUATING TIME-SCALING MODELS OF FORMANT TRAJECTORIES

Following application of SSANOVA and GAMMs as described in §4.4.2 and §4.4.3, the formant trajectories in voiceless vs. voiced coda conditions for each vowel within each city's vowel inventory were compared under the three time-scaling models: time-normalization, proportional-scaling with left-alignment, and proportional-scaling with right-alignment. Three evaluations per formant, per vowel were conducted to determine the best-fitting time-scaling model in each case:

1. *SSANOVA visual-comparison evaluation method:* the three formant-trajectory time-scaling models' SSANOVA outputs were compared, and the best-fitting model in each case was determined via observation of maximal overlap of pre-voiced and pre-voiceless splines.
2. *GAMMs visual-comparison evaluation method:* the best-fitting model was selected via observation of the pre-voiceless condition difference smooth which most closely matched the pre-voiced reference level, determined by least distance from and/or overlap with the zero-line.

3. *GAMMs AIC-score evaluation method:* AIC scores were calculated for each time-scaling model, and the lowest-scoring model among the three was determined to be the best fit.

The results from the three single-formant evaluations described above were compiled in order to determine the best-fitting model per vowel. For each vowel, every evaluation which was in agreement for a particular time-scaling model was summed; each model thus received a score between 0 to 6, being the maximum achievable from three evaluations multiplied across two formants. In the simplest and strongest of cases, all three evaluation methods were unanimous across both F1 and F2 for a given vowel, yielding a score of 6, and the indicated time-scaling model was thus determined to be the best-fitting model for that vowel. In cases of non-unanimity, a score of 5 is fairly strong as it derives from agreement across both formants via two evaluation methods, with a single formant in the third method giving a different result. A score of 4 is achieved via one of two paths: agreement across both formants via one evaluation method, with the other two methods giving different results across F1 and F2, having one positive "hit" for the majority selection; or, two evaluation methods yielding full agreement across both formants, with the third method obtaining a different result for *both* formants. Under this rubric, I consider a score of 4 to be quite weak and probably unreliable. Lesser scores, i.e. below 4, indicate substantial disagreement between methods (all three methods must disagree on at least one formant), disallowing a confident determination of best-fitting time-scaling model for that vowel; as such, any score of 3 or less was simply rated "no preference".

Due to the novelty of this evaluation method, I include two examples from the Winnipeg dataset to illustrate different evaluation scenarios: /ɔj/ with the strongest possible evaluation score of 6, and /ɑ/ with a weak score of 4.

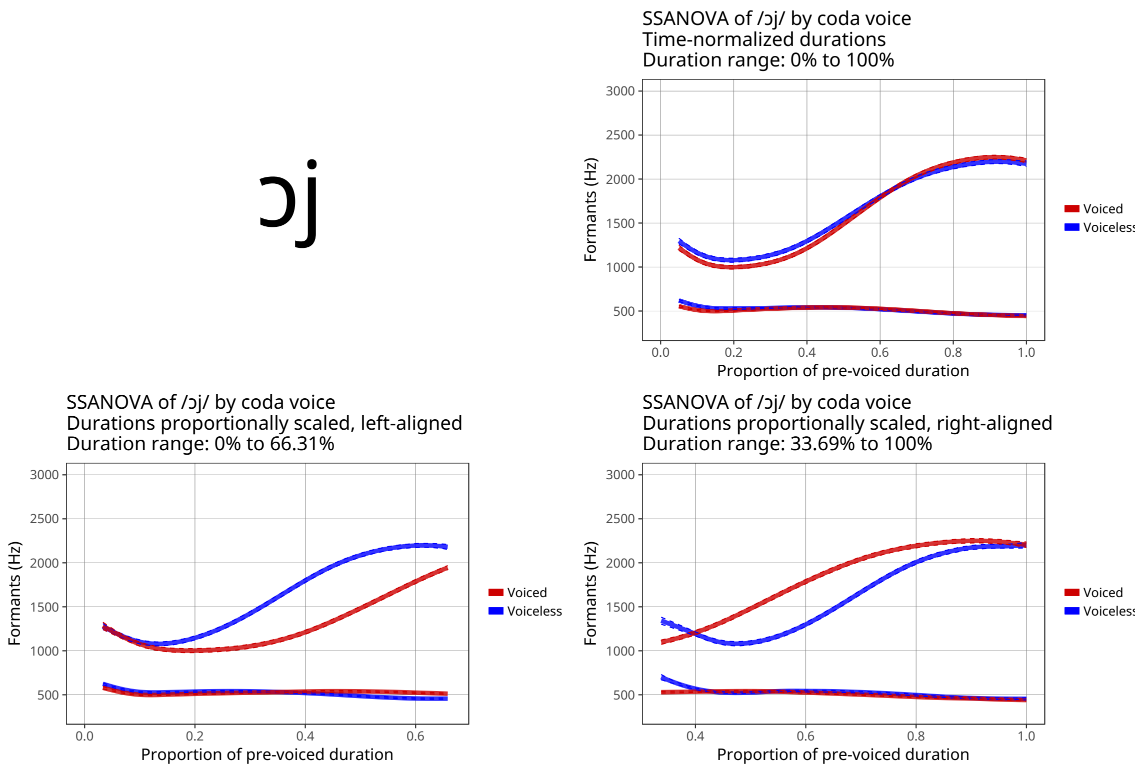


Figure 4.30 SSANOVA time-scaling models for Winnipeg /ɔj/

Taking /ɔj/ first, the initial step in the evaluation involves visual comparison of the three SSANOVA formant-trajectory time-scaling models for F1 and F2, presented in Figure 4.30. Visual comparison of the three models shows that both the F1 and F2 trajectories adhere more closely to each other in the time-normalized comparison (upper right) than under either of the proportionally-scaled models (lower two graphs; this is more blatantly obvious for F2, but F1 is also handled slightly better by time-normalization). The superior handling of both formants by time-normalization provides 2 points towards the score of this model for /ɔj/.

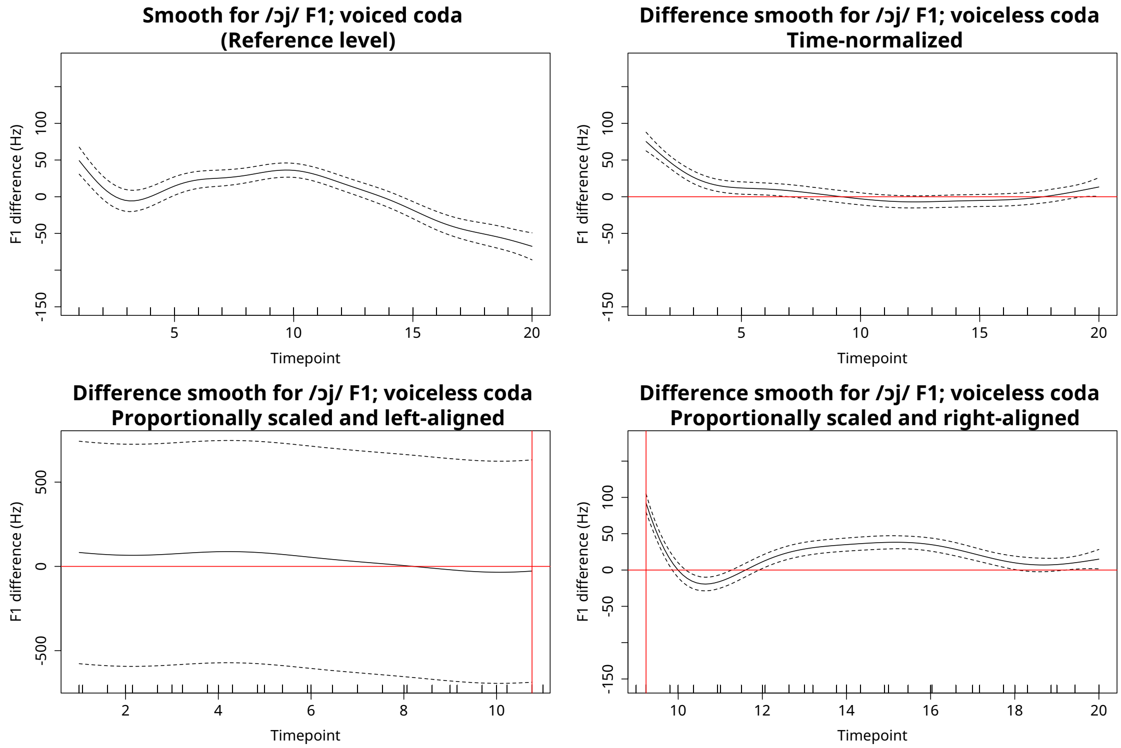


Figure 4.31 GAMMs time-scaling models for Winnipeg /ɔj/ F1

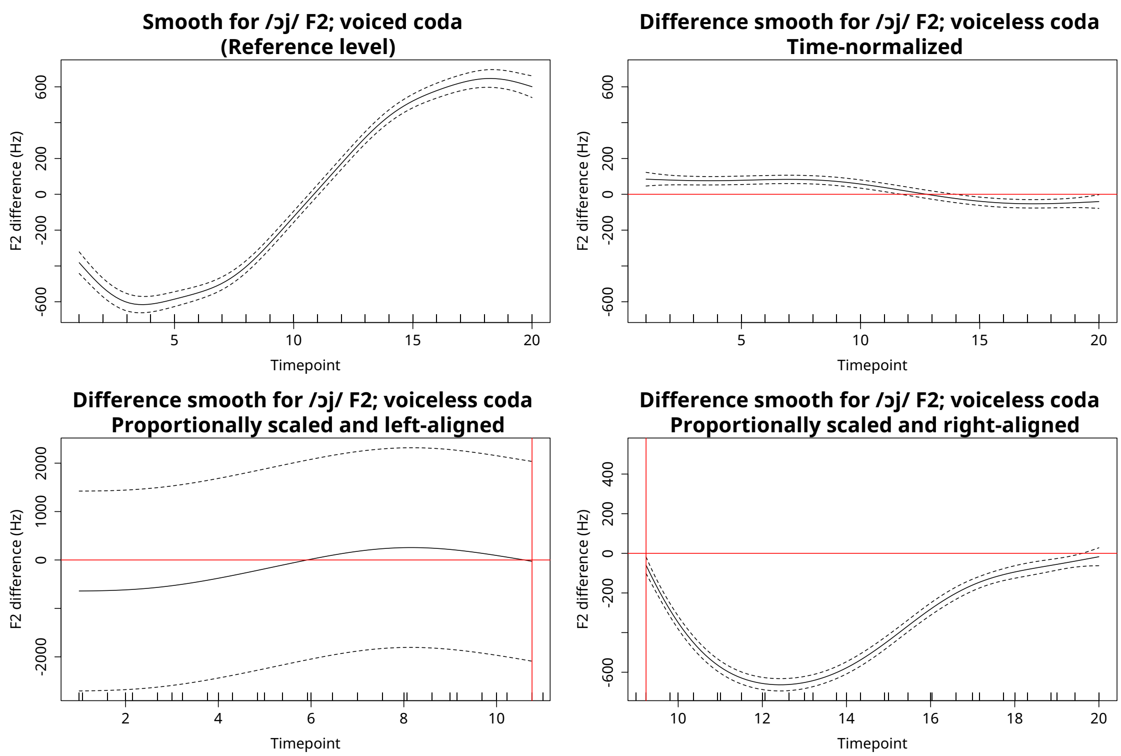


Figure 4.32 GAMMs time-scaling models for Winnipeg /ɔj/ F2

The second step is to visually compare the time-scaling GAMMs models for each formant, as shown via difference smooths for F1 (Figure 4.31) and F2 (Figure 4.32). Under GAMMs, difference smooths which adhere closely to the y-axis zero-line are considered non-significantly different from the reference, which is the pre-voiced condition in this case. In both Figure 4.31 and Figure 4.32, the time-normalized difference smooths (upper right) both remain closer to the zero-line. Additionally, when compared with the proportionally-scaled, left-alignment difference smooths (lower left), the confidence intervals (dashed lines) are quite narrow. These comparisons provide another 2 points towards the score for the time-normalization model for /ɔj/.

Finally, the AIC results for the GAMMs comparisons need to be considered. Under the GAMMs code implemented here, raw AIC scores were not provided as output, only the differences between the AIC scores of any two models under comparison. For F1, the time-normalization model was lower than the proportionally-scaled left-alignment model by an AIC of –208.27, and was also lower than the proportionally-scaled right-alignment model by an AIC of –202.91. For F2, the same pattern was observed, with time-normalization achieving lower AIC scores in each case, by –223.24 points vs. proportionally-scaled left-alignment, and by –287.51 points vs. proportionally-scaled right-alignment. This provides another 2 points to the overall score for time-normalization for /ɔj/, and a final score of 6, indicative of unanimous agreement across all three methods for both formants.

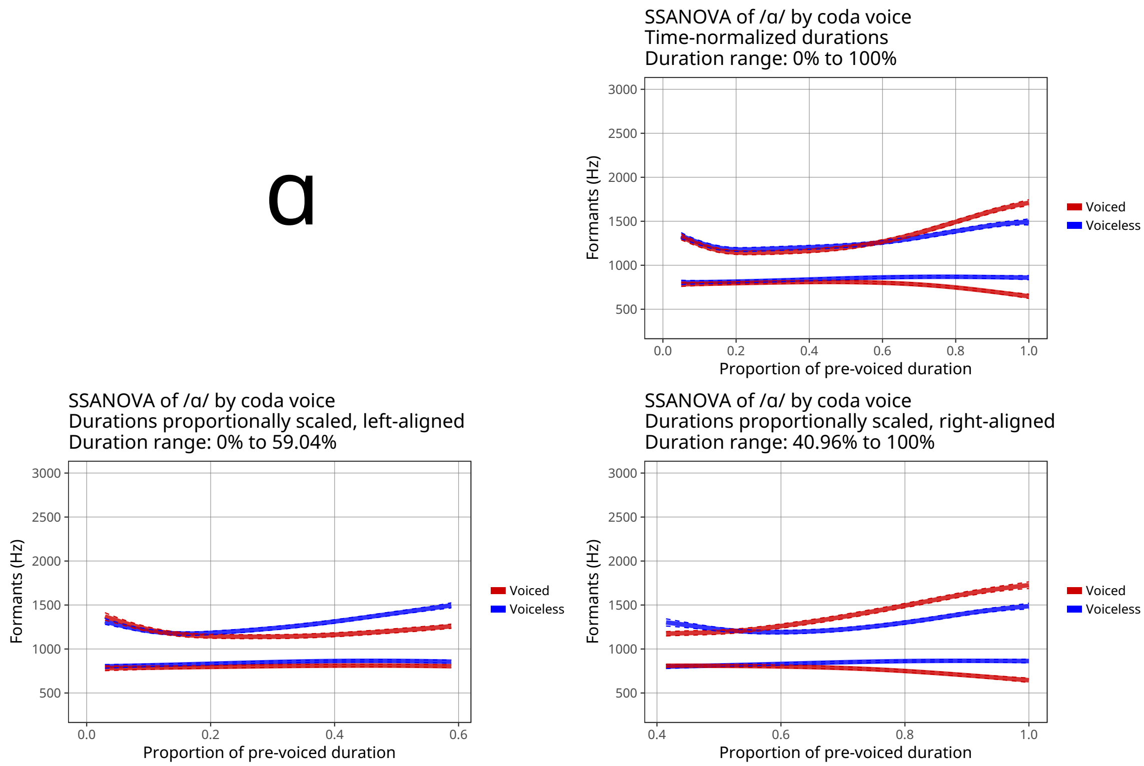


Figure 4.33 SSANOVA time-scaling models for Winnipeg /ɑ/

To illustrate another scenario where the evaluation achieves less consensus, the vowel /ɑ/ is examined. SSANOVA comparisons of F1 and F2 are shown first, in Figure 4.33. F1 fares best under the proportionally-scaled left-alignment model (lower left), as the other two models have divergent F1 trajectories towards the right edge, while F2 favours time-normalized (upper right), albeit somewhat less definitively; all models exhibit divergent F2 trajectories, but time-normalization is simply less divergent than the other two models. The proportionally-scaled left-alignment, and time-normalization models thus each score 1 point.

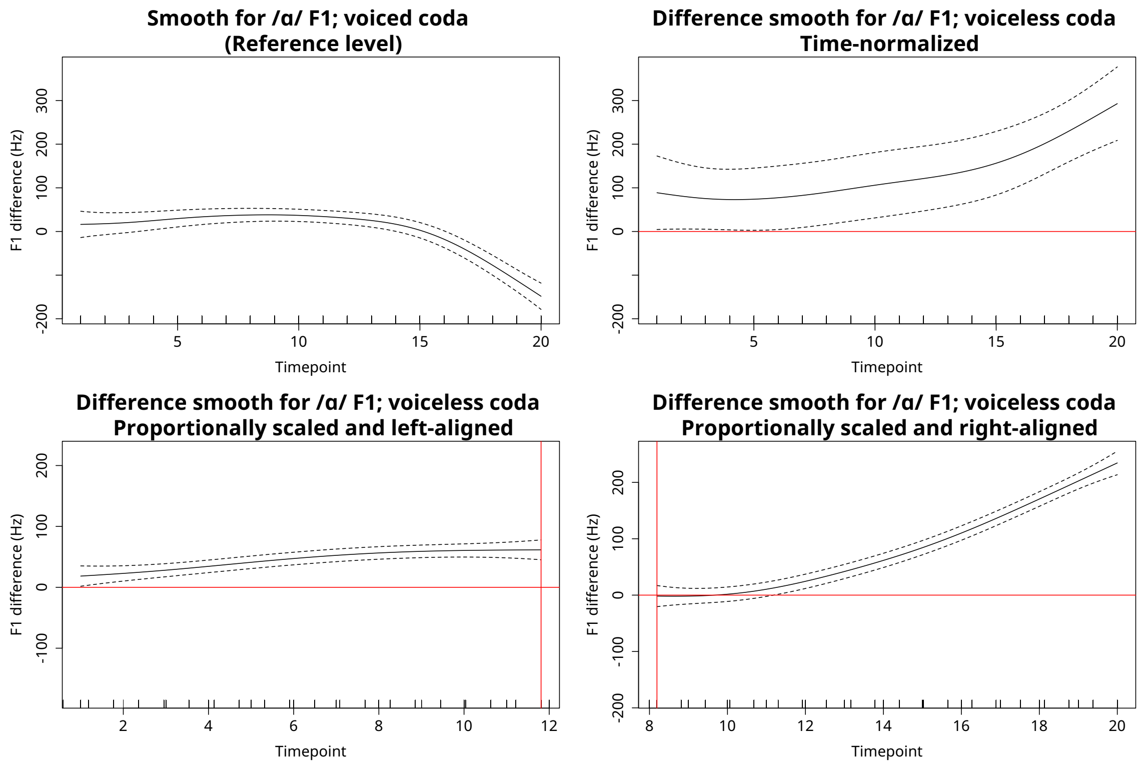


Figure 4.34 GAMMs time-scaling models for Winnipeg /ɑ/ F1

GAMMs visual comparisons of F1 (Figure 4.34) and F2 (Figure 4.35) offer similarly mixed results. For F1, although none of the difference smooths completely overlap the zero-line, the proportionally-scaled left-alignment model (lower left) diverges the least.

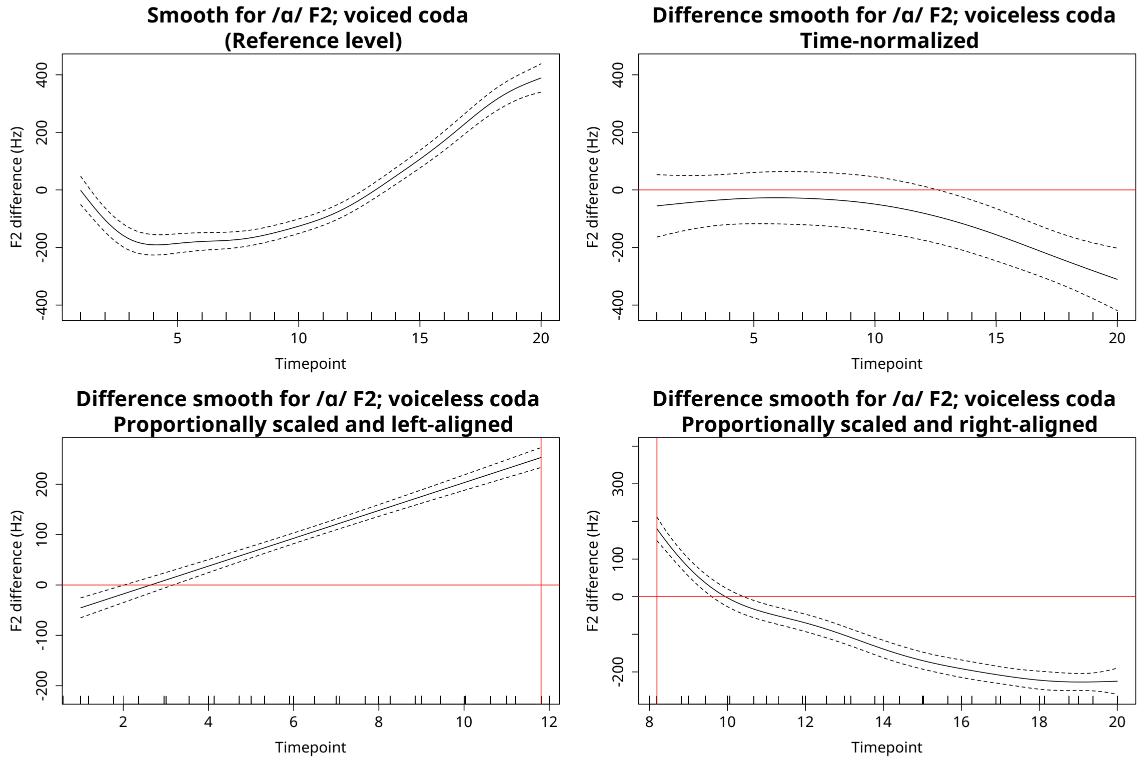


Figure 4.35 GAMMs time-scaling models for Winnipeg /ɑ/ F2

For F2, only the time-normalization model exhibits any degree of overlap with the zero-line, and it is actually fairly substantial (albeit with relatively wide confidence intervals), lasting for approximately two-thirds of the duration. The proportionally-scaled left-alignment and time-normalization models again achieve 1 point each via this method.

Finally, GAMMs AIC comparisons are considered for /ɑ/, offering a resolution. For F1, time-normalization is superior to the other two models, by –66.6 points vs. proportionally-scaled left-aligned and by –150.02 points vs. proportionally-scaled right-aligned. For F2 as well, time-normalization is superior, by –49.57 points vs. proportionally-scaled left-alignment and by –30.89 points vs. proportionally-scaled right-alignment. This adds 2 points to time-normalization's overall score, bringing it to 4 vs. proportionally-scaled left-alignment's 2 points, indicating that it is the overall superior model for /ɑ/.

The various time-scaling models may be taken to imply specific articulatory implementations of PVVA—compression vs. truncation—as described in §4.4.1. Under this view, time-normalization indicates a compression regime, while proportional-scaling indicates truncation, with left- vs. right-alignment indicating the respective non-truncated portion of the articulation for the abbreviated PVVA form. From this perspective, a "no preference" determination for a given vowel indicates that the different available implementations of PVVA are indistinguishable on a practical basis for that vowel, via the methodology described above. Considering that "flat" formant trajectories, which are the acoustic outcomes of relatively static articulatory trajectories, will remain relatively similar whether abbreviated via either compression or truncation, it makes sense that a large number of vowels should fall into the "no preference" category, or obtain a weak score of 4, as not all vowel articulations are especially dynamic in character. This does not necessarily indicate that particular processes of PVVA implementation do not occur for such relatively static vowels, but merely that we may not be able to distinguish between them via comparisons of formant trajectories, which are what the SSANOVA and GAMMs tests are meant to achieve. It can therefore be expected that diphthongal vowels—whether canonical diphthongs such as /aj, aw/ or monophthongs with a substantial off-glide such as /e, o/—should be more readily categorized under the procedure described above than more truly monophthongal vowels, such as /ɛ, ʌ/; the former are known to exhibit a substantial amount of change in their formant trajectories (to varying degrees) over the course of their durations, whereas the latter typically do not.

The vowel inventories of each dataset (Winnipeg, Denver, Madison) were categorized according to the procedure described above, yielding the best-fitting time-scaling model per vowel. The results of these evaluations are displayed in Figure 4.36 through Figure 4.38 as vowel plots in a (mean) F1–F2 grid[32]. The position of each vowel is noted by a marker whose shape and colour identify the selected model. The relative size of the vowel and associated marker indicates the reliability of the determination: the largest size indicates an evaluation score of 6, i.e. a unanimous determination, intermediate size indicates a 5, i.e. a fairly strong determination, and the smallest size indicates either a 4 (weak determination) or "no preference" rating, the latter also distinguished by a distinctive shape and colour (pink diamond).

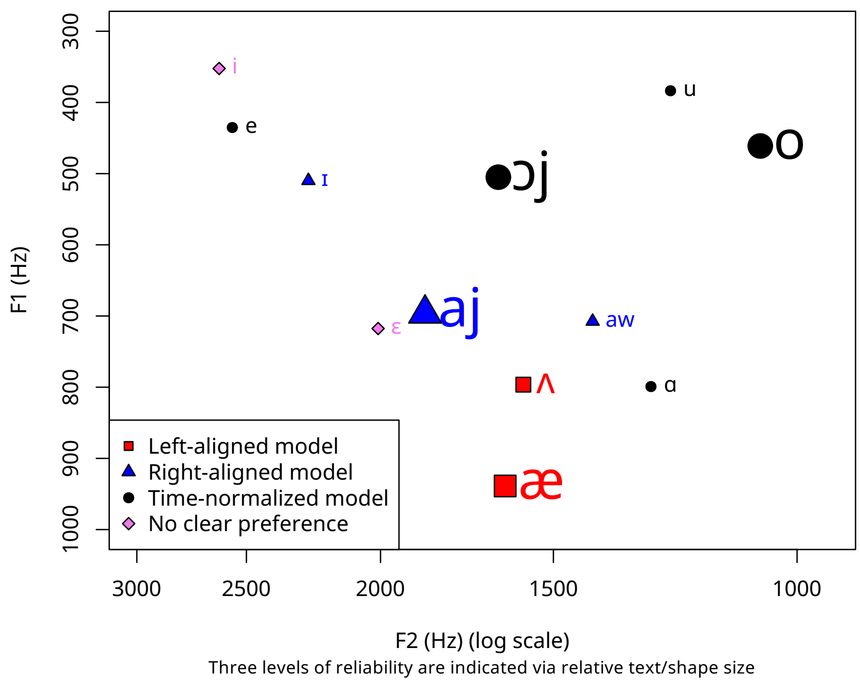


Figure 4.36 Best-fit time-scaling model per vowel: Winnipeg

---

[32] Vowel positions in these plots are identical to those presented in §4.1, with the addition of the diphthongs whose positions are also calculated via simple mean F1 and F2 values, i.e. ignoring their dynamic qualities.

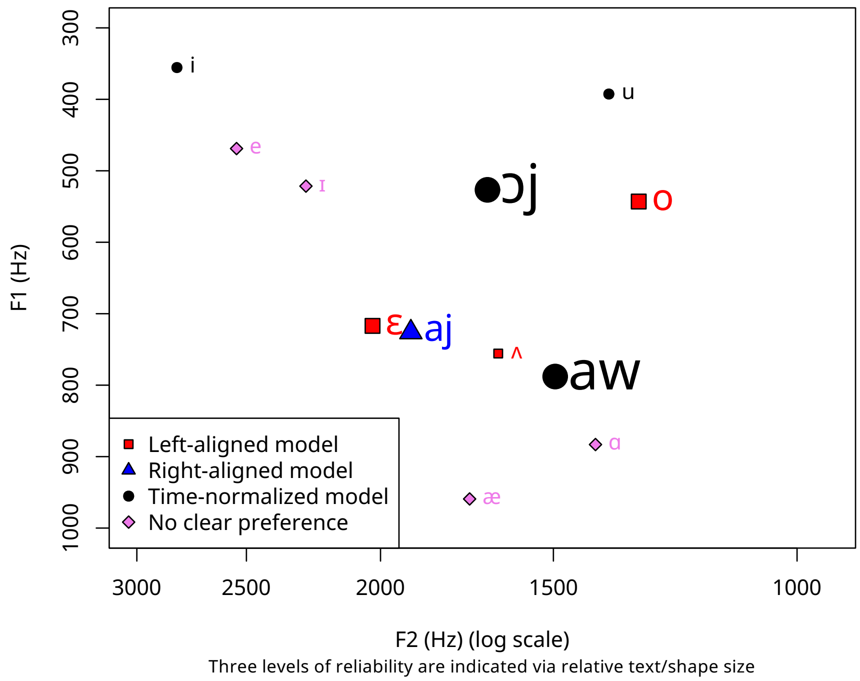


Figure 4.37 Best-fit time-scaling model per vowel: Denver

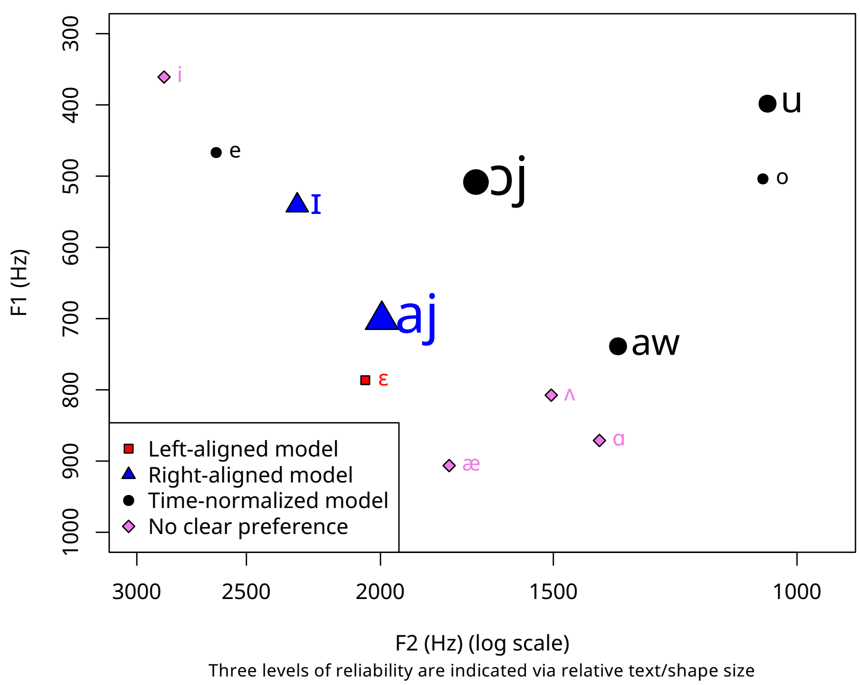


Figure 4.38 Best-fit time-scaling model per vowel: Madison

Speaking generally, the results obtained across all three cities indicate that diphthongs achieve more reliable determinations than monophthongs, which is in concordance with the expectation that diphthongal articulations should be more readily categorized for a particular time-scaling/PVVA model[33]. Focusing on the canonical diphthongs /aj, aw, ɔj/, the results here are in broad agreement with the results discussed in §4.3, where /aj, ɔj/ were found to be highly similar across all three cities, with /aw/ being divergent in Winnipeg alone. Here, /ɔj/ achieves the highest level of overall consistency, with a unanimous determination for the *time-normalization* or *compression* model in every city. /aj/ achieves slightly less consistency, only in that the determination level in Denver is somewhat weaker than in the other two cities; nonetheless, the particular model is identical across all three, *proportional-scaling with right-alignment* or *truncation of onset*. Finally, /aw/ achieves a strong determination in Madison and a unanimous determination in Denver for *time-normalization (compression)*, while in Winnipeg a weak determination is made for *proportional-scaling with right-alignment (truncation of onset)*. It is notable that none of the canonical diphthongs were determined to have *proportional-scaling with left-alignment (truncation of offset)* as a best-fitting model, although I don't think that anything more general can be said about this finding based on the limited data here; a larger examination of diphthongs in other dialects and languages would be required in order to contextualize that result.

[33] This may imply that the high-reliability (score of 5 or 6) determinations made for certain monophthongs—such as Winnipeg /æ, o/ in Figure 4.29—may reflect a more diphthong-like articulation; these two vowels do exhibit a degree of diphthongality in their formant trajectories, although it may also be said that they are not alone in this regard among the monophthongs (see Appendix B for the full set of Winnipeg SSANOVA comparisons, and Appendix E for the GAMMs comparisons).

Chapter 5 Discussion

The method of analysis which I have developed for comparing durationally-distinct formant trajectories in this dissertation, and the results from its application to the specific datasets involved, raise some important questions with regard to the phonology of CR, and of PVVA. One of the most important lines of thought concerns the phonological implications of the PVVA modeling results. I will discuss this generally first, followed by a specific implementation within the framework of Articulatory Phonology, a phonological approach which is particularly suitable to my methodology. This implementation will involve consideration of several important topics, including: how diphthongs are structured internally; how voicing differences are handled; and what motivates the choice of different abbreviation methods (compression vs. truncation). Once these have been described, I will also consider some of the implications my particular approach has for the investigation of diphthongs. Two particular points will be addressed. The first concerns the question of flattened, weakened, or monophthongized diphthongs: are these amenable to the same kind of approach taken here to handle raising patterns? The second point concerns what the model predicts more generally. Given the specifics of the Articulatory Phonology model which I describe, what types of patterns might we expect to find concerning diphthong realizations in the face of PVVA, and which patterns might we expect to be absent?

Another topic of discussion concerns the perceived qualitative differences between CR diphthong allophones. Under the approach I have taken, CR should be seen as essentially an outcome of the abbreviation process (PVVA)—with some important dialect-specific adjustments as described in the Articulatory Phonology model, to follow. If CR is basically a dialect-specific PVVA pattern, why is it not typically described as such? A related topic, which I will discuss following this, concerns the transcription of diphthongs. Selecting notational forms necessitates consideration of both the specific values obtained from close phonetic analysis, as well as the general patterns observed cross-dialectally.

In the following sections I will address these questions and discuss some potential approaches to answering them; areas worthy of more detailed investigation in the future, beyond what I have achieved here, and for which answers are less forthcoming at present, are also discussed.

5.1 THE PHONOLOGICAL IMPLICATIONS OF ABBREVIATION MODELLING

As described in Chapter 4, to analyze differences in vocalic allophones across PVVA contexts (i.e. before voiced vs. voiceless codas), SSANOVA and GAMMs comparisons of PVVA-differentiated vowel formant trajectories were conducted, revealing three distinct patterns across different vowels and dialects. The first broad pattern involves vowels whose formant trajectories are largely similar across different-duration allophones; under statistical comparisons, these are most similar under a *time-normalization* treatment. The second broad pattern involves vowels whose formant trajectories in shorter-duration allophones resemble only a duration-matched portion of the formant trajectories of the longer-duration allophones; these are most similar under a *proportionally-scaled* comparison. The latter pattern has two sub-patterns, one where trajectory-matching involves the leftmost portion of the longer-duration allophone, the other where it involves the rightmost portion. Although the terms *time-normalization* and *proportional-scaling* refer to methods of data-scaling when conducting curvilinear statistical comparisons, I suggested that the three observed patterns may be described as available *time-scaling models* of PVVA, a description with phonological implications.

Cross-comparison of SSANOVA and GAMMs results were used to determine the preferred model on a per-vowel, per-dialect basis; preferred-model determinations, where a conclusive result could be obtained, are aggregated in Table 5.1.

Table 5.1 Cross-dialect comparison of preferred PVVA time-scaling model by vowel

City	Time-normalized model	Proportionally-scaled model Left-aligned	Proportionally-scaled model Right-aligned
Winnipeg	**e,** ɑ, **u, o, ɔj**	æ, **ʌ**	**ɪ, aj**, aw
Denver	i, **u, aw**, **ɔj**	**ɛ**, o, **ʌ**	**aj**
Madison	**e, o, u, aw**, **ɔj**	**ɛ**	**ɪ, aj**

Commonalities across 2+ dialects in boldface

A fair number of commonalities were found across multiple (two or more) dialects, as indicated with boldface type. Excluding results at the weakest level of determination, preferred time-scaling models, as established via total or near-unanimity of the determination (see §4.4.4 for details on the rubric used), are displayed Table 5.2.

Table 5.2 Cross-dialect comparison of preferred PVVA time-scaling model by vowel, high reliability determinations only (score of 5 or 6)

City	Time-normalized model	Proportionally-scaled model Left-aligned	Proportionally-scaled model Right-aligned
Winnipeg	o, **ɔj**	æ, ʌ	**aj**
Denver	**aw**, **ɔj**	ɛ, o	**aj**
Madison	u, **aw**, **ɔj**	—	ɪ, **aj**

Commonalities across 2+ dialects in boldface

In comparison with Table 5.1, the similarities across multiple dialects in terms of high-reliability determinations in Table 5.2 are entirely associated with the diphthongs.[34] I believe that this points to the particular suitability of this method for application to diphthongal data, indicating that monophthongal data is not an especially useful application for this method, or monophthongs are not as strongly differentiated according to the three time-scaling models in comparison with diphthongs, or, perhaps most likely, both of these things.

Underlying my description of the different-duration formant trajectory patterns as *time-scaling models* is an assumption that a phonological process (or processes) is involved. To be explicit, I begin with the assumption that durationally-distinct phonetic realizations of vowels which occur in PVVA and non-PVVA contexts relate to the same underlying phoneme, one whose fundamental structure is more closely related to *one* of the allophonic forms.[35] For example, the vowel in both *beat* and *bead* is assumed to be phonologically identical, despite having significantly different phonetic durations in each word (and all such minimal pairs involving the same distinction of coda voicing; and, likewise for every vowel). This assumption alone, however, is not sufficient to establish which allophone, that having either the longer or the shorter duration, is the "elsewhere" allophone, and thus more representative of the underlying structure of the phoneme. Given only the *time-normalization* model, that is where formant trajectory patterns appear similar

[34] Although monophthong /o/ achieves high-reliability ratings in two dialects, in Winnipeg and Denver, they are non-identical determinations: *time-normalized* vs. *proportionally-scaled & left-aligned*, respectively.

[35] While this may be a contentious point for some, it relates to the particular phonological framework, Articulatory Phonology, which I adopt for discussing a phonological model of PVVA (below); therefore, I ask such readers to suspend judgement for the time being should they disagree with this characterization of the allophonic relationship.

in both short- and long-duration allophones, i.e. irrespective of any durational differences, a reasonable argument could be made for either form. This could involve, on the one hand, a process of durational *compression* of the longer form to create the shorter, and on the other, a process of durational *expansion* of the shorter form to create the longer version. However, there are two facts which conspire to suggest that the longer-duration allophone *is* the base form, and that the shorter version is derived from it through a process of *abbreviation*, hence the term *pre-voiceless vowel abbreviation*.

The first observation concerns vowels in open syllables[36], i.e. with no following coda. For the most part, durational differences between vowels in open syllables vs. vowels before voiced codas are substantially smaller in comparison to the durational differences between vowels before voiced vs. voiceless codas (see Table 4.2). For several vowels, mean durations are nearly identical between open syllables vs. voiced codas, but in all cases vowels before voiceless codas exhibit (typically by a large margin) the shortest durations among all syllable types (see Figure 4.5 through Figure 4.7). That is, the phonetic duration of vowels in open syllables is more similar to, and in certain cases nearly overlapping with, the duration of vowels before voiced codas, than either of these is to the duration of vowels before voiceless codas. As a vowel occurring *before a voiceless coda* describes a more restricted environment than the combination of *in open syllables and before voiced codas*, it seems more plausible to posit that the form occurring in the latter environment represents the basic form of the phoneme, rather than the converse.

---

[36] I should emphasize that as the data for this dissertation focuses on monosyllables, all vowels tokens in open syllables examined are also in word-final position; durational comparisons in other positions within a word may differ.

The second observation concerns the formant trajectory comparison pattern leading to the *proportional-scaling* model, and its two attendant sub-models of *left-* vs. *right-alignment*. For vowels adhering to this model, the shorter-duration allophone's formant trajectories correspond to only a *portion* of the trajectories of the longer-duration allophone. In order to posit that the shorter-duration forms are more basic, and that the longer-duration allophones are derived from them, some phonological information is required to be added, as the longer-duration trajectories are non-linear, and therefore cannot be completely predicted from the shorter-duration trajectories. Furthermore, the added phonological information, i.e. the non-corresponding parts of the longer-duration trajectories, are different *for every vowel*. A phonological process that adds *different information* to each phoneme which undergoes it on an ad hoc basis does not appear to be a tenable one. On the other hand, positing that the longer-duration allophone is more basic, the shorter-duration forms can all be derived quite simply by specifying what proportion of the longer-duration form to utilize, and beginning at which "edge", i.e. left- vs. right-alignment, and thus eliminating or *truncating* the remaining portion, a more plausible and coherent phonological process. Considered along with the previous observations regarding open syllables, this suggests that the longer-duration vocalic allophones should be taken as the basic phonemic form, from which the shorter-duration forms are derived. This further implies that in time-normalization, *compression* rather than *expansion* is the active mechanism for achieving PVVA.

Adopting this position that the shorter-duration allophones are derived *from* the longer-duration allophones indicates that the difference in vowel duration before voiced vs. voiceless codas is an abbrevatory process, i.e. PVVA. The two basic time-scaling models, time-normalization versus proportional-scaling, therefore imply two distinct phonological processes which are utilized to achieve PVVA: *compression*, and *truncation*, respectively. The task of describing these processes in a particular model of phonology, namely Articulatory Phonology, will be taken up in the following section.

5.2 Articulatory Phonology and PVVA

Articulatory Phonology (AP; Browman & Goldstein 1986, 1989, 1992, 1995; inter alia) offers a principled method for describing phonology as the alignment and coordination of articulatory *gestures*, defined as "organised patterns of movement within oral, laryngeal and nasal articulatory systems," (Browman & Goldstein, 1986:225). I believe this approach to phonological modeling offers the potential to capture both methods for achieving PVVA, compression and truncation, in a substantive and coherent manner. The fundamental units in AP, known as *gestures*, are sub-phonemic in nature, and their temporal relationships can be manipulated in a variety of complex ways. AP is thus well-suited to a phonological process such as PVVA which produces allophonic patterns of systematically altered phonetic duration. While a fully-realized phonological model of PVVA and/or CR under AP (or any other theory), as noted earlier, is beyond the scope of this dissertation, in the following subsections I offer a brief sketch of some of the most important aspects of such a model, and those which are able to be most clearly explicated at present, while also outlining those areas that remain to be addressed comprehensively.

5.2.1 THE FRAMEWORK OF ARTICULATORY PHONOLOGY

Articulatory Phonology is a theory of phonology that derives differences in surface articulatory and acoustic phonetic output from differences in the presence, type of, and coordination between articulatory *gestures*, defined (above) as "organised patterns of movement within oral, laryngeal and nasal articulatory systems," (Browman & Goldstein, 1986:225), which are assumed to be the basic units of phonology. AP was first outlined by Browman & Goldstein, and has since been elaborated by other researchers (see e.g. Browman & Goldstein, 1989, 1992, 1995; Saltzman & Munhall, 1989; Byrd, 1996, 2003; Gafos, 2002; Gafos & Goldstein, 2012; inter alia). In AP, the (traditional) phonological segment is viewed as being composed of a constellation of gestures involving one or more of the various articulatory organs. One of the insights of AP is that adjacent segments may have their component gestures coordinated in ways that are difficult to capture within other frameworks, but which lead naturally to descriptions of observable phonetic and phonological phenomena (arguably, more so than other theories). As such, the temporal or dynamic component of gestures is a defining feature of AP: "As actions, gestures have some intrinsic time associated with them — they are characterisations of movements through space and over time" (Browman & Goldstein 1989:201).

AP builds upon and incorporates many aspects of a related articulation-based model of speech production termed *task dynamics* (Saltzman 1985; Saltzman & Kelso 1987; Saltzman & Munhall 1989). The task-dynamics framework defines a set of *model articulators* and *tract variables*. The *articulators* include: the upper and lower lips; the jaw; the tongue tip; the tongue body or dorsum; the velum; and, the glottis. These are employed to generate constrictions within the vocal tract as specified by the various *tract variables*, which include: lip protrusion (LP); lip aperture (LA, the distance between the lips); horizontal lip motion (LH); independent upper and lower vertical lip motion (ULV and LLV); jaw angle (JA); lower teeth height (LTH); tongue tip/body (dorsum) constriction location/degree (TTCL, TTCD, TBCL, TBCD); radial and angular positioning of the tongue body (TBR and TBA); the velic aperture (VEL); and the glottal aperture (GLO). In the task-dynamics framework as adopted in AP, a given gesture comprises a specific configuration or *constellation* of tract-variable specifications, which indicate the location and degree of the constriction(s) within the vocal tract. There are a range of specifications which are also associated with the tract variables, describing various types of constrictions which may occur such as *close, wide, narrow*, etc. which have the effect of differentiating different manners of articulation (albeit not specifically described as such), e.g. fricatives vs. stops. Some of the most commonly utilized tract variables are illustrated in Figure 5.1.

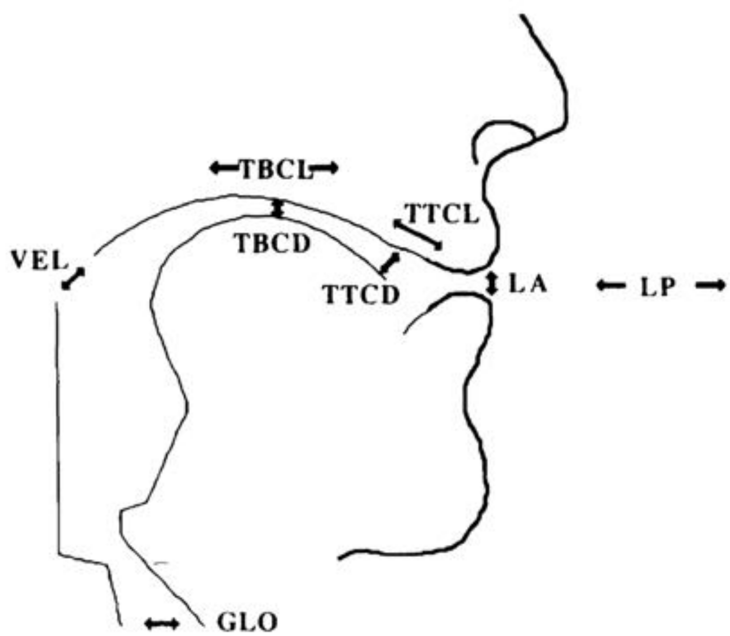


Figure 5.1 Tract variables and contributing articulators of computational model (Browman & Goldstein 1989:207)

The series of gestures comprising an utterance are illustrated in AP via a *gestural score* which indicates their relative order and timing, as well as the type of constriction to occur, e.g. *close, wide, narrow*. An example of a traditional AP-style gestural score is provided in Figure 5.2 for the word *palm*. The "box notation" used in this type of gestural score utilizes dashed-line boxes indicating the period of time over which a given gesture is activated. Constriction types are indicated both with prose notation, e.g. "wide", as well as by dynamic solid-line paths indicating the relative movement of the various articulators, e.g. the rising line indicating the state of the velic aperture (VEL) at the right edge of the upper portion of the score.

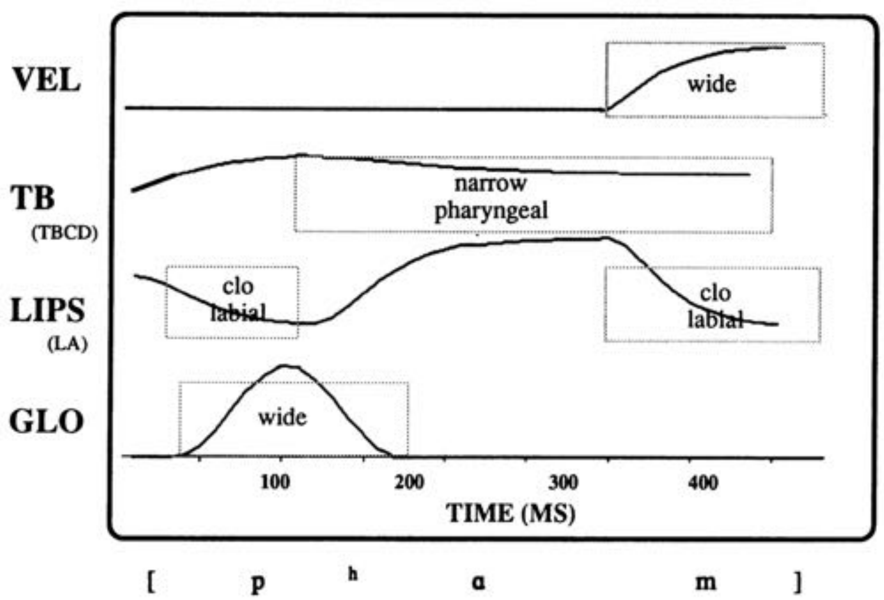


Figure 5.2 Gestural score for *palm* [pɑm] using box notation, with model generated tract variable motions added (Browman & Goldstein 1989:201)

As understood in AP, a single gesture is a dynamic action which takes place over an interval of time, and gestural scores specify the relative arrangement of multiple gestures over time. Although gestural scores incorporate temporality, they lack a key component necessary for a complete model of the dynamic patterns of gestures, as they do not in themselves indicate precise sequential timing relationships between gestures. For instance, in Figure 5.2 above, although the LIPS closure and GLO widening gestures are non-simultaneous, with the latter initiating slightly after the former has already begun, the mechanism by which this particular timing relationship is established or maintained is not specified.

In order to offer a means of dealing with such issues, Gafos (2002, 2003) proposed an account of gestural coordination in AP using the concept of gestural *landmarks*. Figure 5.3 illustrates the set of gestural landmarks proposed by Gafos as they occur in sequence over the "life" of a gesture: *onset*, *target*, *c-center* (simply, "the mid-point of the gestural plateau," Gafos 2002:271), *release*, and *release-offset*. The gestural *plateau* comprises the region from the *target* to the *release*, inclusive, and marks the portion of the gesture during which its particular constriction is fully achieved and maintained.

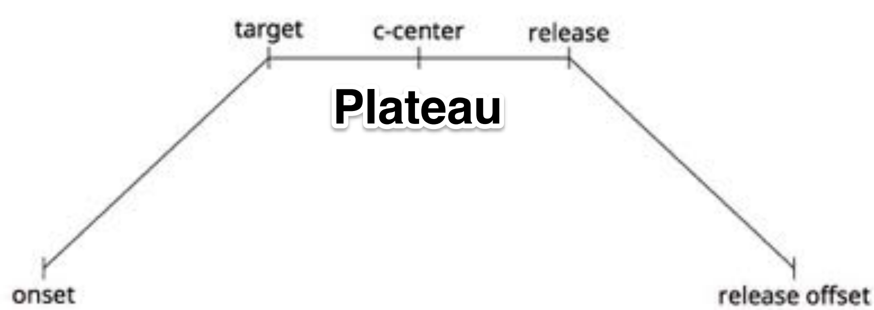


Figure 5.3 Gestural landmarks and the gestural plateau (based on Gafos 2002)

Coordination between gestures in Gafos' system is achieved through the *gestural coordination relation*, "a relation between two gestures stating that a specified landmark (within the temporal structure) of one gesture is synchronous with a specified landmark of another gesture," (Gafos 2002:278). Gafos implements various possible gestural coordination relations as Optimality-Theoretic constraints by adapting the existing constraint ALIGN to cover gestures ('G') and gestural landmarks; the basic schema for all such relations is shown in (1):

1) ALIGN (G^1, landmark^1, G^2, landmark^2): Align landmark^1 of G^1 to landmark^2 of G^2
Landmark^i takes values from the set {ONSET, TARGET, C-CENTER, RELEASE}[37]

The language of Optimality Theory (OT) provides a useful way to discuss gestural coordination, as OT is well-suited to handling competing pressures within a theoretical model. The combination of various articulators and gestures which exist within any AP gestural score will inevitably produce competition between full and careful articulation of each gesture, ease of articulation by the speaker, functional acoustic output, and accurate perceptibility of the output by the listener, among other factors. OT also allows interaction between constraints such as (1) above, and others concerned with other considerations in predicting in phonological structure, e.g. markedness and faithfulness. Although such constraint interactions will not be addressed in this dissertation, they would inevitably form part of a complete model of English phonology, including PVVA and other processes which could affect or constrain its occurrence in various contexts.

A basic division between *consonant* and *vowel* gestures is implicit in much of the AP literature, and stated explicitly by Goldstein, Byrd & Saltzman (2006) who argue that it is fundamental to the composition of syllables: "the internal structure of syllables results from different ways of coordinating gestures of these basic types [i.e. consonants and vowels]"

---

[37] This set of landmarks should also include *release-offset* as well, as indicated by Gafos' own VV COORD constraint in (2) below; in other words, all gestural landmarks are potential points of alignment.

(p. 229). Gafos (2002) describes four distinct cross-segmental *coordination relations*, which cover all of the possible permutations of two-segment sequences of consonants (C) and vowels (V), presented here in (2):

2) CV COORD(INATION): ALIGN (C, C-CENTER, V, ONSET)

The onset of a V is coordinated to the c-center of a preceding C

VC COORD: ALIGN (V, RELEASE, C, TARGET)

The target of a C is coordinated to the release of a preceding V

CC COORD: ALIGN (C^1, RELEASE, C^2, TARGET)

The target of a C is coordinated to the release of a preceding C

VV COORD: ALIGN (V^1, ONSET, V^2, RELEASE-OFFSET)

The onset of a V is coordinated to the release-offset of a preceding V

Schematics of the four[38] individual gestural coordination relations in (2) are illustrated in Figure 5.4[39]; the landmarks which are coordinated under each gestural coordination relation are labelled with distinct colours relating the landmark to its particular gesture, and the points of coordination are marked with shaded circles.

---

[38] Gafos suggests that, among the proposed gesture coordination relations, only CV COORD and VC COORD are universal, while CC COORD and VV COORD are not universal, which offers a means of expressing systematic variation across languages/dialects; I won't delve into the argumentation on this point, as it doesn't appear to play a role in the development of the PVVA model to follow.

[39] Although structurally distinct in terms of coordinated landmarks, the CV COORD and VC COORD constraints/relations appear to produce identical patterns of coordination as displayed in Figure 5.4; differences between these two relations only become apparent when multiple Cs are present. As onset consonants and complex onsets/codas are not discussed in this dissertation, these particular differences between CV COORD and VC COORD will not be examined further. CC COORD, at least in its English-language version as provided here, also patterns in the same way.

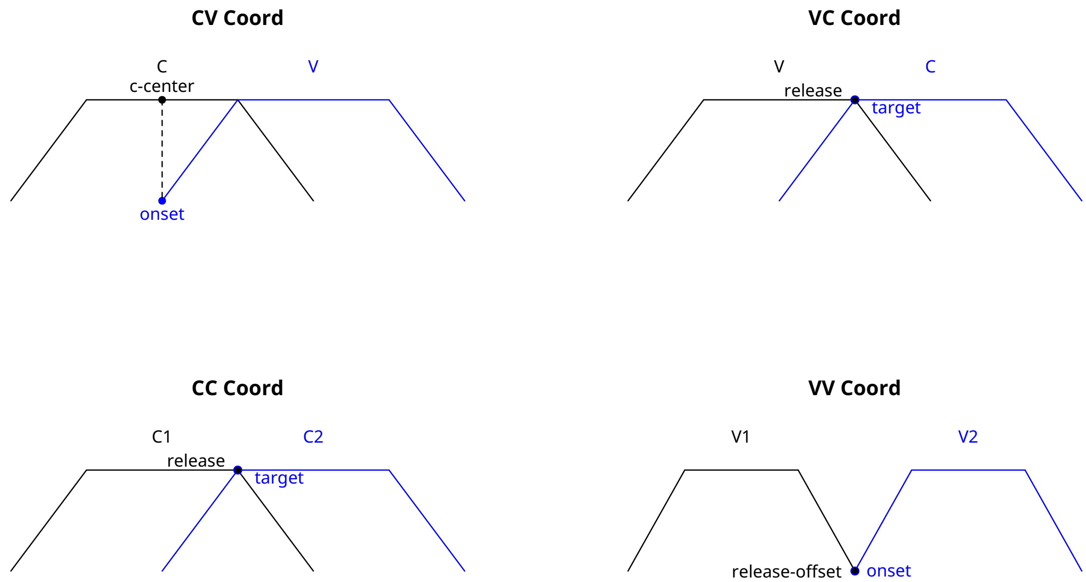


Figure 5.4 Gesture coordination relations (based on Gafos 2002)

Gafos' coordination relations address cross-segment gestural coordination, but a comprehensive account of segment-*internal* gestural coordination is lacking both in Gafos' coordination system and, it should be noted, most discussions of more traditional AP. It seems fairly obvious in viewing a gestural score, such as in Figure 5.2, that segment-internal gestural coordination will impact cross-segmental coordination, due to the fact that each segment's component gestures do not necessarily begin or conclude at the same time. The topic of such internal coordination is therefore of considerable importance for a model of PVVA and will be taken up in several of the subsections to follow.

As a final point regarding AP theory, I should emphasize that the *gesture* in AP is fundamentally an abstract, phonological construct, despite its grounding in phonetic output, i.e. the association of each gesture with a physical articulator. The full AP computational system involves several distinct components, with the gestural score serving as input to the task-dynamics model which resolves the score into various articulatory trajectories, as depicted in Figure 5.5.

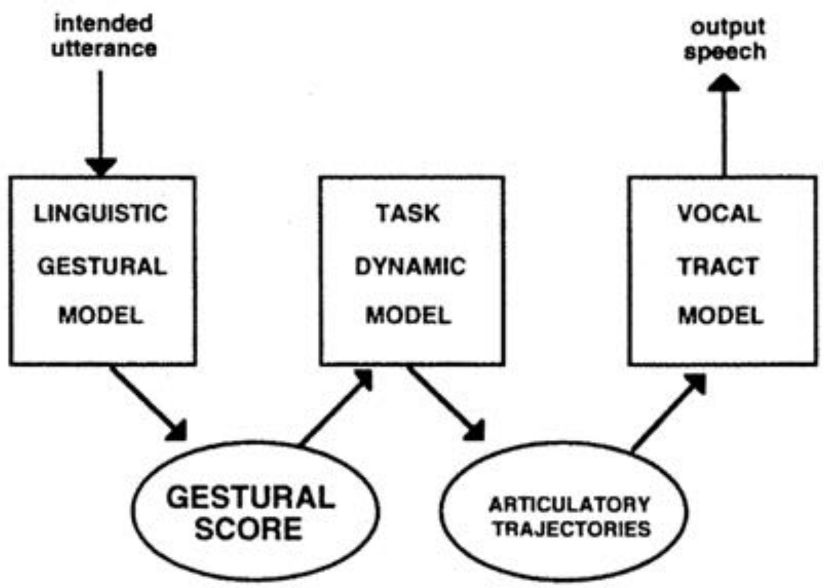


Figure 5.5 Computational system for generating speech using dynamically-defined articulatory gestures (Browman & Goldstein, 1995:55)

As the gestural score occupies a fairly abstract level within the computational system, there is not a one-to-one match between *phonological* gestures and *phonetic* articulations deriving from those gestures. As depicted in Figure 5.5, the gestural score feeds into the task dynamic model, which interprets the score to generate the requisite articulatory trajectories, and which are themselves constrained in various ways by the vocal tract model. Nonetheless, AP gestures are grounded in phonetics by being explicitly linked to physical articulators. As will be seen in the sections to follow, a sufficient model of PVVA in AP will rely in large part on the interpretation of (phonological) gestural overlap, including overlapping gestures of the same articulator. At the phonological level, such overlap may specify that an articulator be in two different positions simultaneously, a physical impossibility and hence something which need to be resolved in favour of a single phonetic outcome.

5.2.2 DIPHTHONGALITY AND METHODS OF ABBREVIATION

Most of this dissertation has been concerned with the characteristics of diphthongs, especially the CR diphthongs /aj, aw/ but also the third English diphthong /ɔj/. As such, the interpretation of diphthongality in AP is a requirement for a comprehensive AP model of PVVA and CR. This is a topic which is sorely under-researched in the AP literature, however; the only relatively active body of research in this regard concerns the Romanian diphthongs /ea, oa/ (Marin 2005, 2007a, 2007b, 2014). These diphthongs alternate phonetically between diphthongal and monophthongal realizations, i.e. [ea~e] and [oa~o], respectively. For this reason, Marin (2007a) proposes an AP account of Romanian diphthongality wherein two same-articulator vocalic gestures (e.g. TB) are coordinated to occur *synchronously* rather than *sequentially*. Different phonetic outcomes result from greater or lesser degrees of gestural overlap between the gestures at the phonological level, related to speech rate and stress effects, and the favouring of one gesture over the other. In Marin's account, when complete gestural overlap occurs between the component gestures of Romanian diphthongs due to rapidity of speech, priority is given to the initial gesture, such that the phonetic outcomes are equivalent to those of phonologically monophthongal /e, o/; in intermediate cases, the two gestures are "blended" to greater or lesser degrees.

Due to the rather different natures of the languages involved, I will not adopt specific aspects of Marin's account of Romanian here. However, the phonetic consequences of gestural *overlap* and *prioritization* or *favouring* are very much relevant issues. Based upon the results discussed in §4.4, I make the following assumptions regarding the gestural makeup of English diphthongs, with respect to the three dialects investigated in this dissertation:

Assumptions for modeling diphthongality within the AP framework

1. diphthongs are produced via multiple (i.e. at least two) distinct vocalic gestures
2. diphthongal gestures may not overlap completely, such that a phonetic monophthong is produced[40]
3. portions of diphthongal gestures may be truncated under certain conditions, producing an abbreviated phonetic output whose formant trajectories only match a portion of a longer-duration allophone produced in a different context
4. a sequence of diphthongal gestures may be compressed durationally under certain conditions, producing an abbreviated phonetic output which exhibits similar formant trajectories to those of a longer-duration allophone produced in a different context

Gafos' landmark system provides a coordination relation, VV COORD for the purpose of coordinating sequences of vowels; however, Gafos (2002) explicitly notes that this pattern is reserved for syllable-timed languages such as Italian where "the timing of vowels is unaffected by the length of the intervening consonantal period," (p. 325). In order to discuss English off-gliding diphthongs, I therefore adopt Gafos' VC COORD for this

[40] This statement clearly does *not* apply to certain dialects where "flattened" diphthongs may occur.

purpose, as glides are typically considered as consonantal in nature. The structure of the gestures of the diphthong /aj/ under this coordination relation are illustrated in Figure 5.6.

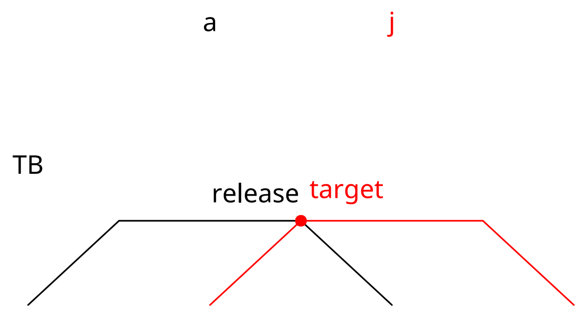


Figure 5.6 Diphthongal gesture coordination for /aj/

Under the assumption that the pre-voiced allophone represents the basic gestural structure of the phoneme, I assume that the coordination pattern shown here corresponds to the phonetic output of pre-voiced /aj/, whose mean formant trajectories are shown in Figure 5.7; although results from the Winnipeg dataset are illustrated here, the outcomes for /aj/ are fairly similar in the other two dialects examined, as discussed in §4.4.

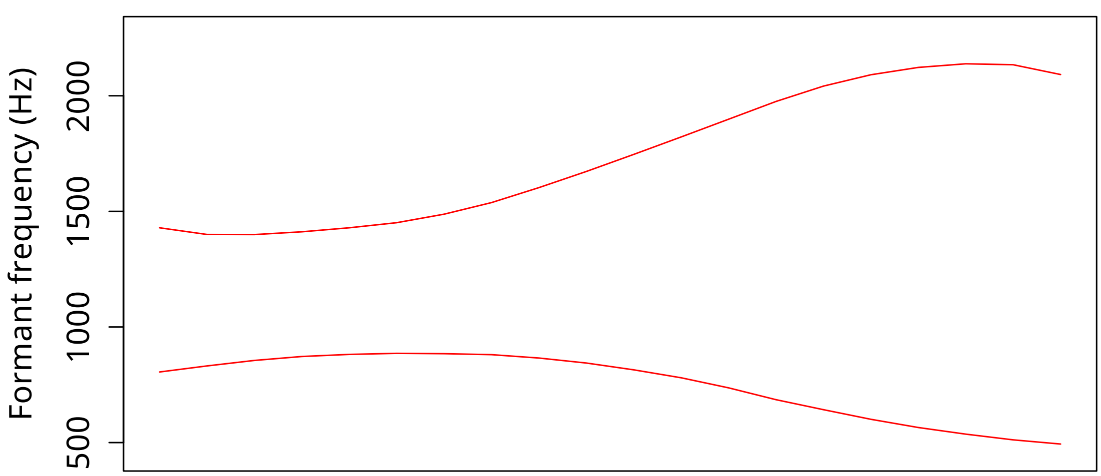


Figure 5.7 Mean formant trajectories of pre-voiced /aj/, Winnipeg

The plateaus of the two diphthongal gestures may be seen to correspond to particular regions over the duration of /aj/, as shown in the following schematic in Figure 5.8.

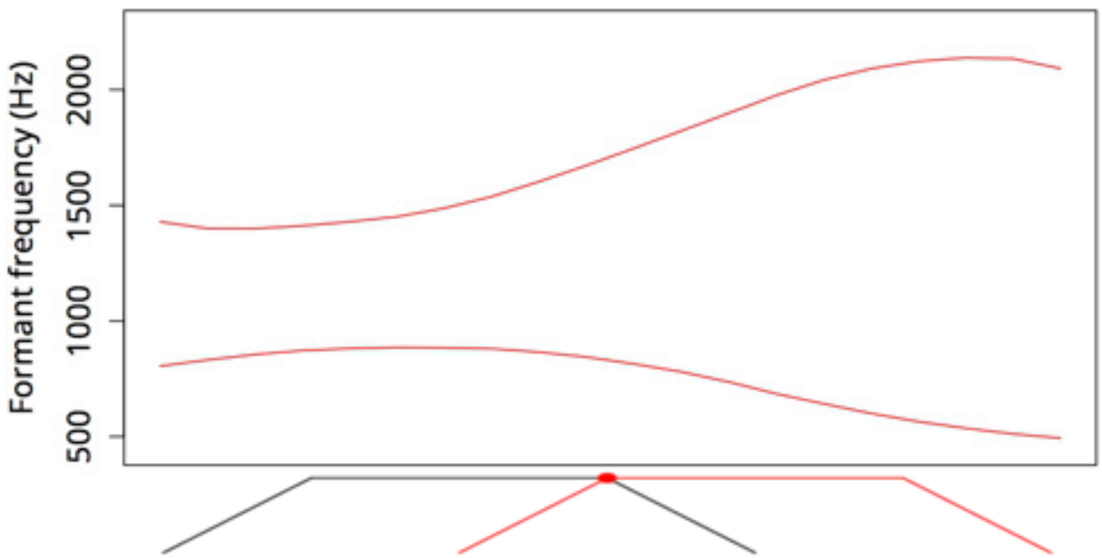


Figure 5.8 Formant trajectories (top) and gestural score (bottom) for /aj/

Clearly, there is not a one-to-one correspondence between gestures and regions of phonetic output. Nevertheless, a relationship between the two is indicated in this visual comparison. On the left, a lengthy steady state in the formant trajectories corresponds to the (black) nuclear gesture. In the central region the formant trajectories are in a dynamic state of transition, corresponding to the region around the point of coordination between the nuclear (black) and glide (red) gestures in the gestural score. It's clear that this period of transition in the formant trajectories extends further to the right in the figure than does the region of gestural overlap; in addition, the steady state in the formant trajectories towards the right edge (vowel offset) is much briefer than that which is associated with the nucleus. It seems reasonable that there would be a degree of lag involved in moving from one same-articulator gesture to another, which would not necessarily be explicitly represented in the gestural score itself, but which would account for the disparity between the left and right halves of the formant trajectories, despite the parallel patterns of both halves of the gestural score.

The fact that different phonetic outcomes in terms of formant trajectories occur for /aj/ in pre-voiceless context suggests a different pattern of coordination than in the pre-voiced context. Figure 5.9 illustrates the observed output for pre-voiceless /aj/, and a reconfigured pattern of gesture coordination associated with the pre-voiceless context is shown in Figure 5.10.

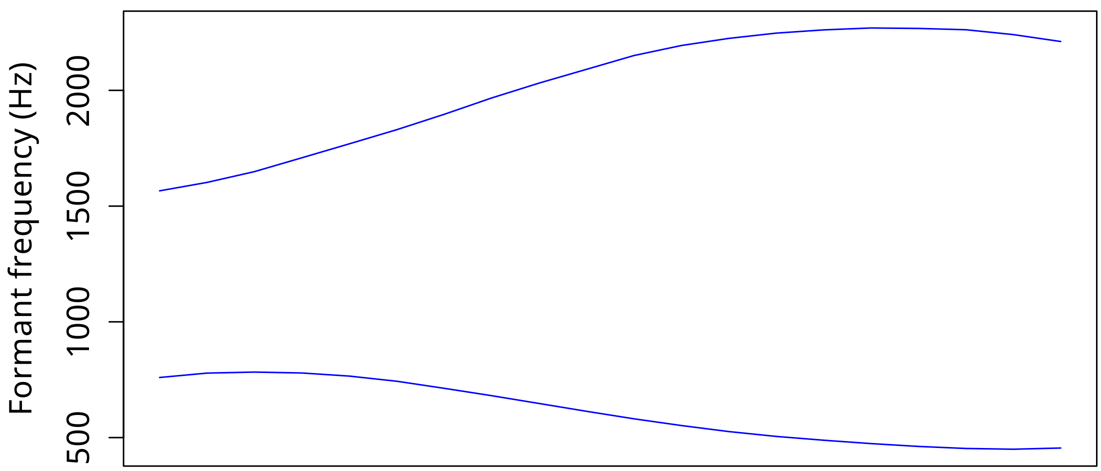


Figure 5.9 Mean formant trajectories of pre-voiceless /aj/, Winnipeg

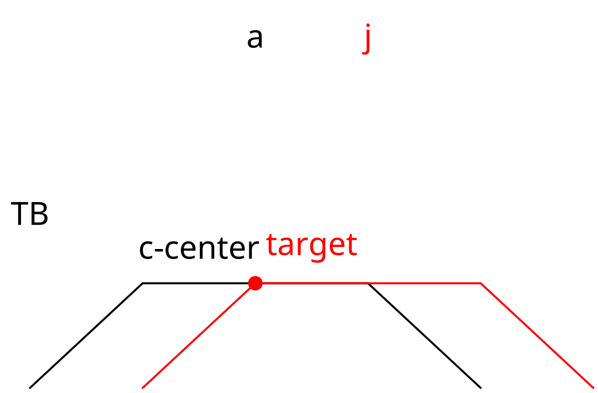


Figure 5.10 Diphthongal gesture coordination for pre-voiceless /aj/

In Figure 5.9, the nuclear steady state of Figure 5.7 is virtually absent. To account for this in AP terms, I propose as shown in Figure 5.10 that the two diphthongal gestures overlap in such a way that reduces the proportion of the diphthong devoted to the nuclear gesture, while still respecting the earlier stated assumption (for the dialects under discussion here) that *diphthongal gestures may not overlap completely, such that a phonetic monophthong is produced*. Coordinating the glide's *target* with the nuclear *c-center* provides the maximal amount of overlap between the two gestures while still maintaining a degree of separation, thus ensuring that a monophthong is not produced as output.

The two gestures overlap for a period of time, in the central portion of the diphthong, which puts competing pressures on the tongue to be in two places simultaneously. Recall that Marin (2007a) suggested that in Romanian, periods of non-complete overlap are resolved by "blending" the gestures to produce an intermediate output. For English diphthongs such as /aj/, I argue that a different strategy is employed: the second, or most recent gesture, in this case the glide, takes precedence from the point of overlap onward. Once the point of gestural coordination is reached, at the c-center of the nucleus, the nuclear gesture ceases its cycle and relinquishes "control" of the tongue, to the glide gesture. It is certainly possible that this particular strategy of "gestural precedence" need not operate in this way in other dialects; I argue only that it applies for the three English dialects under discussion in this dissertation.

In order to incorporate all of the above into a single visual comparison, durational differences between the two allophones must be also considered. Figure 5.11 illustrates the correspondence between both the pre-voiced and pre-voiceless formant trajectories and their respective gestural coordination patterns, incorporating these durational differences by scaling the pre-voiceless allophone's formant trajectories and gestural score according its proportional duration relative to the pre-voiced allophone.

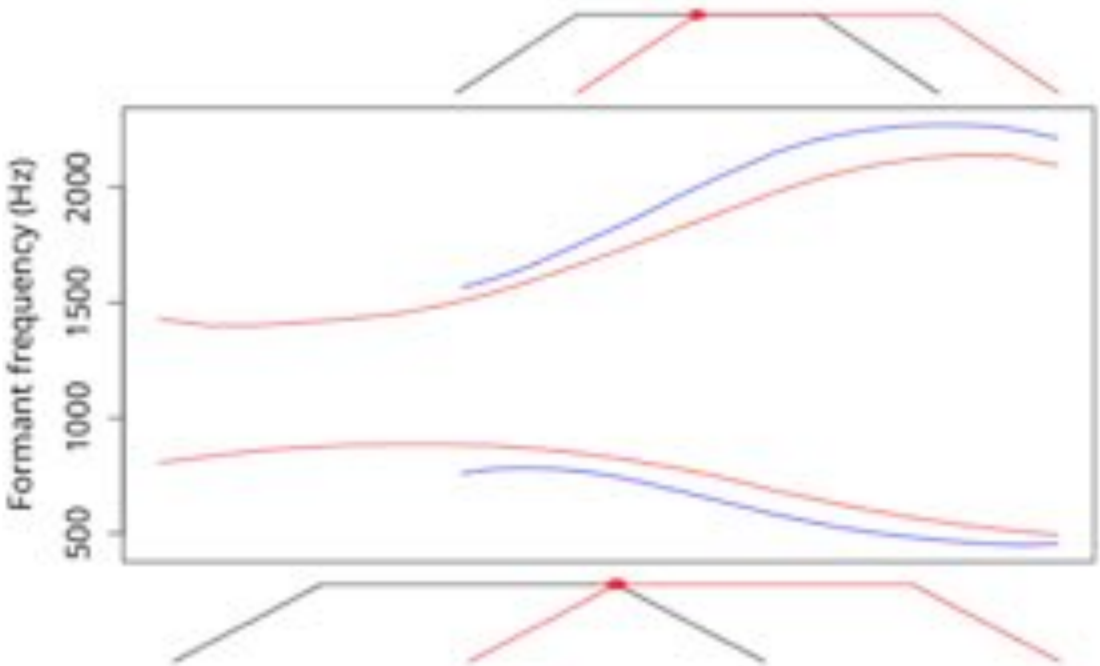
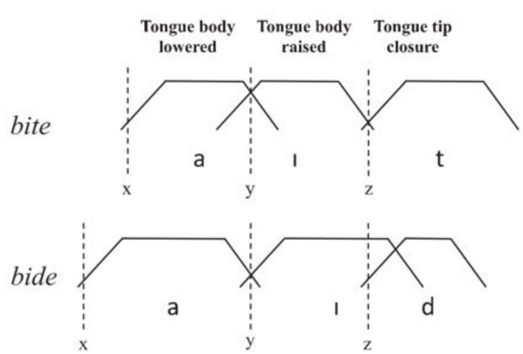


Figure 5.11 Formant trajectories of pre-voiceless (blue) and pre-voiced (red) /aj/, and gestural coordination patterns for pre-voiceless (above) and pre-voiced (below) /aj/[41]

---

[41] Pycha & Dahan (2016) proposed a somewhat similar account of /aj/, offering the following set of gestural scores:



Proposed gestural scores for words with diphthongal vowels before voiceless codas, as in bite [baɪt] (upper panel) and before voiced codas, as in bide [baɪd] (lower panel) (Pycha & Dahan 2016:16)

Their account differs in several important respects from mine, however. First, their subjects are not primarily speakers of /aj/-raising dialects. Second, and more importantly, they did not determine a significant correlation

While I do not assume a strict one-to-one correspondence between AP gestures and particular regions of phonetic output, I think the schematic in Figure 5.11 captures the basic relationships well, and that differences between the two gestural scores systematically predict the observed differences between the phonetic outcomes of /aj/ in each coda voicing context, within a certain margin of flexibility. Recall that there are two sub-patterns for abbreviation-via-truncation: right-alignment, as depicted here, and left-alignment. The latter is exemplified by a few vowels as determined in §4.4.4, including /æ/ in Winnipeg, which had the strongest possible rating. Other less-reliable determinations for truncation with left-alignment include /ʌ/ in Winnipeg, and /ɛ, o/ in Denver. This pattern will be discussed in more detail further below in §5.2.3.

The second major mechanism of abbreviation arises from modeling of formant trajectory time-scaling comparisons as *time-normalization*, interpreted phonologically as durational *compression* of the entire gestural pattern, maintaining its internal structure. Internal temporal speed, i.e. compression, is handled within AP through task-dynamics, under the gestural parameter of *stiffness*, and is probably best explained through the *phase model* of gestural production. In the phase model, "an articulatory gesture is controlled as a 360°, critically damped, mass-spring oscillatory system," (Byrd 1996:140). Cross-gestural timing is handled in by coordinating between specific points or *phase angles* within the 360° cycle of a gesture. The 0° phase angle is the point of gestural initiation; at

---

between vowel duration and coda voicing, unlike the results in the present study. Third and finally, they do not propose principled methods of gestural coordination or reorganization to describe the relationship between their proposed gestural scores: "the co-occurrence of coda voicelessness with a particular gestural score, versus coda voicing with a different gestural score, remain [sic] unexplained," (p. 34).

this point the relevant articulator for a gesture is in its "resting" state (equivalent to Gafos' *onset* landmark). The 360° phase angle is the point of final conclusion of the gesture, when the articulator has again returned to its resting position, and from which point the cyclic motion may be again activated in another gesture (equivalent to Gafos' *release-offset* landmark). Coordination between the 180° phase angle (equivalent to Gafos' *c-center* landmark) of one gesture and the 0° phase angle (onset) of a subsequent gesture is argued to be inherently stable and therefore common cross-linguistically (Goldstein, Byrd & Saltzman 2006).

In the phase model, gestural *stiffness* determines the speed of the cycle. Differences in stiffness affect the duration of gestures, such that "[a] gesture with a lower stiffness (perhaps a vowel) will have an intrinsically longer duration than a gesture with a higher stiffness (perhaps a consonant)," (Byrd & Saltzman 2003:154). Byrd, Kaun, Narayanan & Saltzman (2000) suggested the incorporation of a prosodic tier into AP scores, containing a "π-gesture" which modulates stiffness in associated gestures. A similar approach might be useful for a model of PVVA by having a π-gesture (whether on a *prosodic* tier or elsewhere) coordinated with a compressed (or compressible) vocalic gesture. Coordination of a stiffness-increasing π-gesture with a vocalic gesture (or sequence of gestures) would produce an abbreviated (compressed) vocalic duration, and thus adhere to the PVVA principle, as an alternative to the glottal gesture coordination method which produces abbreviation via truncation.

Research on stiffness modulation, the prosodic tier, and the nature and description of π-gestures appears to be largely concerned with descriptions of prosodic effects, e.g. phrase-final lengthening, rather than processes such as PVVA, which I would not necessarily describe as prosodic in nature. It is thus not well understood at present how π-gestures might be involved at a non-prosodic level, such as coordinating between same-syllable segments as in PVVA contexts. Figure 5.12 illustrates how a stiffness-reducing π-gesture can manipulate (in this case, slowing) the speed of a region comprising multiple gestures, bearing in mind that, as is typical, the application shown here concerns prosodic effects, placing the π-gesture on a separate Prosodic Tier.

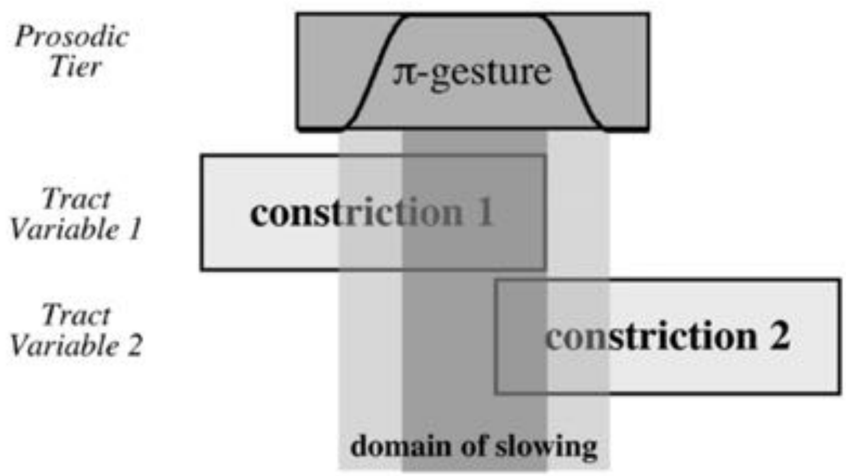


Figure 5.12 A schematic gestural score for two gestures spanning a phrasal boundary instantiated via a π-gesture (Byrd & Saltzman 2003:160)

The diphthong which exhibits a compression pattern across all three dialects in this study is /ɔj/, as illustrated in Figure 5.13.

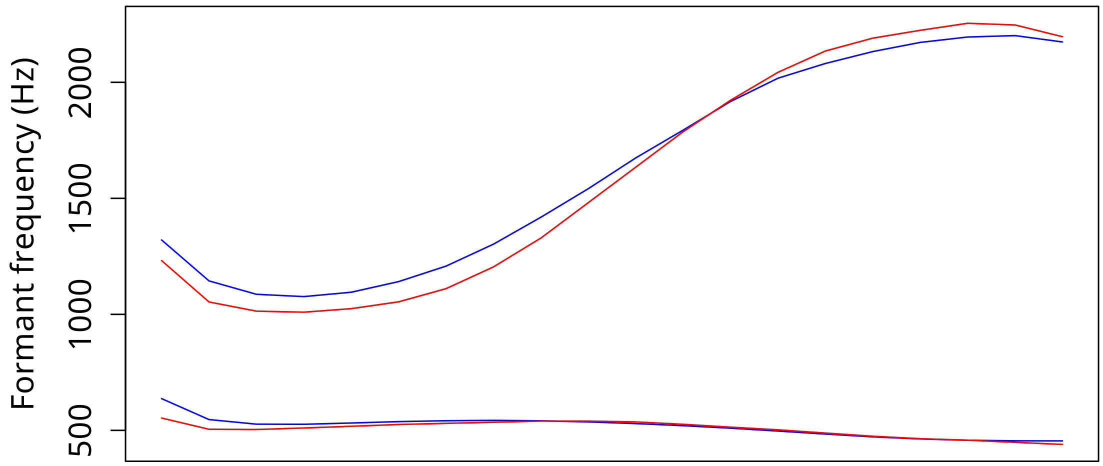


Figure 5.13 Mean formant trajectories of pre-voiced (red) and pre-voiceless (blue) /ɔj/

Because of the similarity of the formant trajectories in each coda voicing context, I posit that the gestural scores for /ɔj/ are likewise similar in each context. This basic gestural score for /ɔj/, in terms of coordinated diphthongal gestures, will appear identical to that of /aj/ as illustrated in Figure 5.6; the only necessary addition is that, in pre-voiceless context, a stiffness-altering π-gesture will be incorporated. The π-gesture in Figure 5.12 *lowers* the stiffness value of any affected gestures within its domain, resulting in a slowing of their articulations and thereby increasing their duration. Conversely, in order to achieve compression, the required π-gesture must *increase* stiffness, thereby producing a *domain of acceleration* which reduces or abbreviates the duration of any gestures affected by it. In the case of PVVA-induced compressed vowels, this domain of acceleration would extend across the entire diphthongal gestural score, resulting in a set of closely matched formant trajectories like those shown in Figure 5.13 for /ɔj/. Note too that the slight differences between the pre-voiced and pre-voiceless allophones of /ɔj/ differ in ways that are compatible with this stiffness-increasing explanation. There is a slight degree of undershoot in the pre-voiceless allophone, most notably for F2, which may result from the more rapid gestural cycle pertaining in this context.

Thus far, I have set up the basic AP concepts required for a PVVA model concerning diphthongal structure, and the manipulation thereof using either gestural overlap or compression in order to achieve abbreviated vocalic allophones. There are two basic components still required for such a model, however. The first of these concerns the motivation for producing abbreviated forms; in other words, what is the rationale or motivation for PVVA, in AP terms? The second component concerns the choice between either of the two available abbreviation mechanisms. Why do all vowels not pattern alike in this regard? Furthermore, why do different sets of vowels pattern differently across dialects? These topics are addressed in the following sections.

5.2.3 PVVA AND GLOTTAL GESTURES

PVVA effects in English, as discussed in §2.1, involve the abbreviation of vowels, including diphthongs, before voiceless codas. Assuming that the discussion in §5.2.2 regarding the composition of, and abbrevatory methods available to, diphthongs in AP is accurate, the question remains: What does an AP model of vowel (and diphthong) abbreviation look like, taking into account coda voicing? The task of accounting for this in AP terms primarily involves the choice of landmark for coordination. In Gafos' VC COORD relation, the post-vocalic consonant's gestural *target* is coordinated to the vowel's gestural *release*. However, this relation does not specify *which* consonantal gesture is involved. This is an issue when considering consonants which differ in voicing quality, and therefore some background information on voicing in AP is first required.

In the standard AP framework, which has largely been applied to languages with solely or predominantly modal phonation types, the state of the glottis is *assumed to be in the position for modal voicing at all times* unless otherwise specified: "We assume that, in speech mode, the larynx is positioned appropriately for voicing unless otherwise instructed," (Browman & Goldstein 1992:157; see also Goldstein & Browman 1986). Under this approach, the only relevant glottal gesture (in a strictly modal-voice language) is glottal abduction or voicelessness, and there is no explicit (phonological) gesture for the production of glottal adduction, or voicing. It should be emphasized that the AP interpretation of voicing is somewhat under-developed relative to other aspects of the theory. The lack of capacity for non-modal phonation within AP has been a noted problem for some time: "Clearly, additional tract variables will have to be defined for non-pulmonic sounds and for words involving contrasting phonation types," (Byrd 2003:90). In addition, the identification of voicelessness with a single gesture associated with the glottis as articulator is certainly an oversimplification. I take the simplistic but pragmatic view here that AP's GLO widening gesture should be associated with any and all states of the vocal tract which induce a state of phonetic voicelessness, without attempting to resolve any further this nonetheless important aspect of the theory.

Adopting the AP view that the only explicit glottal gesture is for the state of voicelessness, Figure 5.14 depicts the articulatory structure of monophthongal *add* /æd/ in a Gafosian framework; although the glottal tier is included in this score, there are no glottal gestures present in *add*, as the modally-voiced status of /d/ is not explicitly specified.

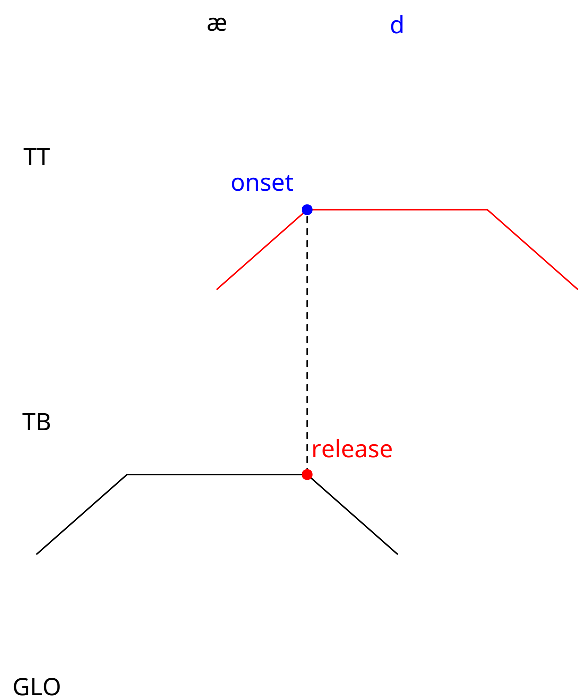


Figure 5.14 Gestural coordination for /æd/

To produce *add*'s minimal pair counterpart *at* /æt/, a voiceless glottal gesture is added to the gestural score, as depicted in Figure 5.15. Here, the two consonantal gestures are coordinated synchronously to the same landmark, the release of the TB gesture of /æ/, in accordance with Gafos' VC COORD gestural relation constraint. Arguably, they may be coordinated in some other pattern, but I see no reason to suggest otherwise at present, and synchronous timing of all of a segment's component gestures appears to be the most parsimonious choice, barring other information to the contrary. Note that in Figure 5.15 and similar following figures, the various consonantal gestures appear on non-adjacent tiers; this is done to emphasize the fact that the gestures comprising any given segment are notionally independent from each other in AP, aside from the indicated coordination linkages.

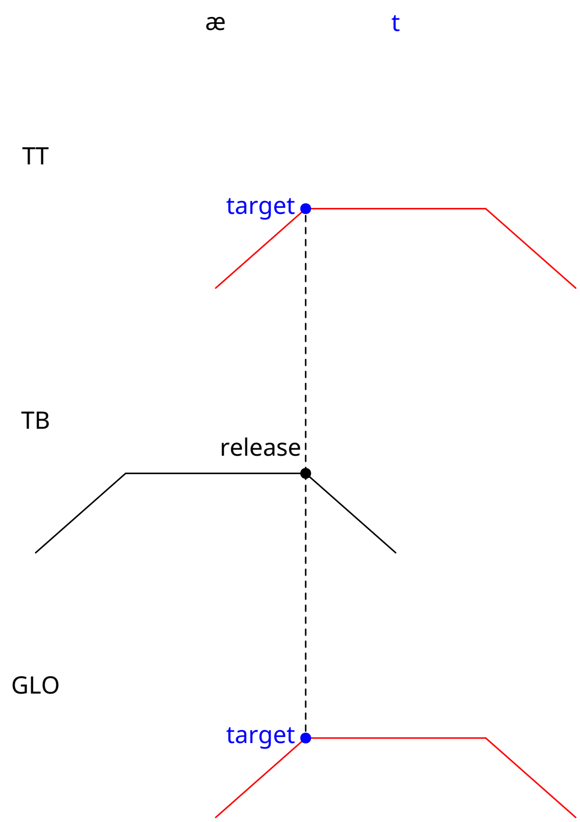


Figure 5.15 Gestural coordination for /æt/, unabbreviated duration

The pattern depicted here does not differ from Figure 5.14 in terms of the portion of the vocalic gesture which is produced before initiation of the consonantal gestures. Because of this, the vowel as depicted in Figure 5.15 will not be durationally abbreviated in comparison to that of Figure 5.14, nor would the acoustic formant trajectories of the vowel produced be expected to differ between the two forms. One way to account for the abbreviation which is observed in the phonetic output in PVVA contexts, is to coordinate the glottal gesture's target to an earlier landmark in the vocalic gesture. In order to avoid complete overlap of the vocalic and consonantal gestures, the only such available landmark is the c-center, as depicted in Figure 5.16.

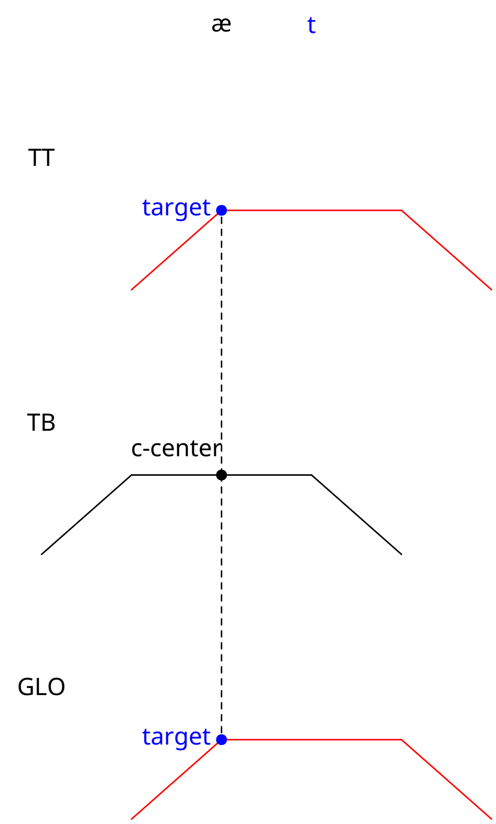


Figure 5.16 Gestural coordination for /æt/, abbreviated duration (offset truncation)

A key element of Figure 5.16 is that the two consonantal gestures maintain coordination with respect to each other, which produces overlap of the two tongues gestures, the TB gesture of /æ/ and the TT gesture of /t/. Although the reconfiguration of Figure 5.16 from the score shown in Figure 5.15 is argued to be driven by the coordination of the GLO gesture with an earlier landmark (the c-center) of the vowel, without also coordinating the TT gesture to this point, there would be no articulator overlap and hence, no abbreviation of the vowel; at best, we could expect a partially-devoiced vowel of similar duration to that of fully-voiced *add* /æd/ (Figure 5.14). Instead, the maintenance of coordination between the two consonantal gestures in Figure 5.16 produces overlap of the two tongue gestures, resulting in an abbreviated vocalic gesture (i.e. PVVA) as achieved via *truncation of the (vocalic) offset*, the pattern which was modelled under formant trajectory comparisons as *proportional-scaling with left-alignment*. Note the similarity between the coordination pattern depicted in the score in Figure 5.16, and that depicting diphthong-internal gestural overlap in Figure 5.10; although two different tiers are involved in the former (TT and TB) vs. a single tier in the latter (TB), the result is the same in that the initial vocalic gesture is truncated via overlap with a subsequent gesture involving the same basic articulator, the tongue.

Figure 5.17 depicts the gestural coordination pattern for non-abbreviated *eyed* /ajd/. This gestural score is a combination of the general diphthong internal gestural structure shown in Figure 5.6, and the (monophthong) vowel-to-voiced-coda structure depicted in Figure 5.14.

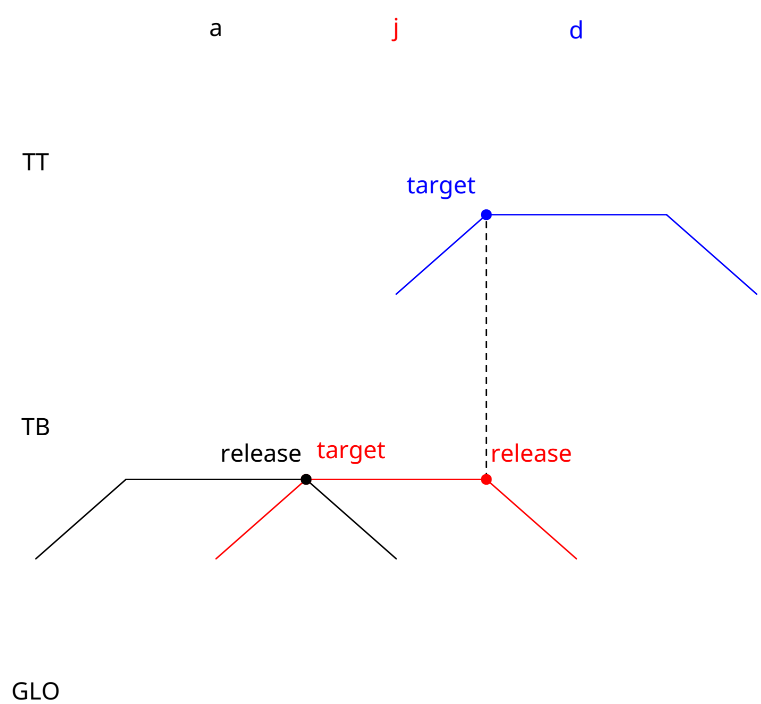


Figure 5.17 Gestural coordination for /ajd/

To produce the counterpart *a'ight* /ajt/, a glottal gesture is added to the configuration, as was the case for the monophthongal *at* /æt/, as shown in Figure 5.18. Here, in keeping with the abbreviation-via-truncation structure depicted for pre-voiceless /aj/ in Figure 5.10, abbreviation occurs not via coordination of the consonantal gestures to an earlier landmark, but rather via re-coordination of the diphthongal off-glide's target to the nuclear c-center.

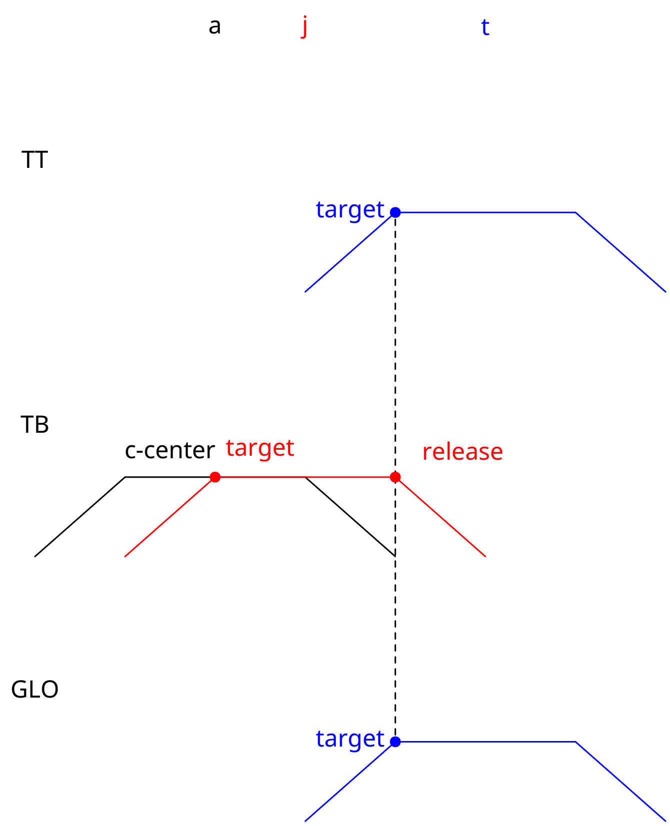


Figure 5.18 Gestural coordination for /ajt/

To achieve abbreviation while avoiding monophthongization, the two diphthongal gestures are maximally overlapped while still maintaining a minimal non-overlapped region. The structure shown in Figure 5.18 describes the (phonological) abbreviation mechanism involving *truncation of the vocalic onset*, and matches the (phonetic) time-scaling model of *proportional-scaling with right-alignment*.

5.2.4 MOTIVATING THE CHOICE BETWEEN ABBREVIATION MECHANISMS

Having discussed the two available major mechanisms for achieving PVVA, compression and truncation, and outlined methods for incorporating both of these into an AP model of PVVA, the final main task for such a model is to describe why certain vowels adhere to one mechanism or the other when undergoing PVVA. This concerns not only the status of different vowels within a single dialect, but also the status of the same vowel across different dialects. The main concern here with regards to the present study involves the diphthong /aw/. In §4.4.4, I noted that the patterning of the two other diphthongs with regards to time-scaling models is identical across the three dialects investigated, whereas /aw/ is divergent in Winnipeg in comparison with Denver and Madison. The groupings among the diphthongs within each dialect in terms of preferred time-scaling model are shown in Table 5.3:

Table 5.3 Cross-dialect comparison of preferred PVVA time-scaling model by diphthong

City	Time-normalized	Proportionally-scaled, right-aligned
Winnipeg	ɔj	aw, aj
Denver	aw, ɔj	aj
Madison	aw, ɔj	aj

As seen here, each dialect presents a two-pronged pattern for diphthongal abbreviation under PVVA, only differing on which pattern /aw/ adheres to. In Winnipeg, i.e. Canadian English, /aw/ patterns like /aj/ under proportional-scaling with right-alignment; under the present discussion, this is handled by the truncation of onset mechanism via diphthong-internal gestural realignment. In Denver and Madison, /aw/ instead patterns like /ɔj/ under the time-normalization model, achieved via the compression mechanism involving a stiffness-increasing π-gesture and its associated domain of acceleration.

What might the impetus be for differentiating between /aj, aw/ vs. /ɔj/ on the one hand, versus differentiating between /aj/ vs. /aw, ɔj/ on the other? The CR pattern involving similar treatment of /aj, aw/ can be argued to arise from their similar (if perhaps not identical—more on this in §5.5.1) nuclei, vs. /ɔj/. For (non-Canadian) dialects which treat /aw, ɔj/ alike, the impetus for doing so may arise from the fact that both involve rounding. I hypothesize that the solution may lie in specifications regarding the coordination of lip-rounding gestures, which would be present in both /aw/ and /ɔj/ in an AP treatment. Unfortunately, there appear to be no systematic studies using AP-based models which have focused directly on rounding gestures, for English or any other language, leaving my suggestions here to be much more speculative in nature than the previous components of the PVVA model discussed thus far.

I will note that although both /aw, ɔj/ may be counted as round vowels, the phonological locus of that rounding is very different, being associated with the off-glide of /aw/ vs. the nucleus of /ɔj/. Perhaps, then, the distinction between Canadian and non-Canadian dialects concerns the coordination of glide-associated rounding gestures on the one hand, and nucleus-associated rounding gestures on the other. For Canadian dialects, these may be treated distinctly, while in non-Canadian dialects they may be handled identically. Both "sets" of dialects handle round-nucleus diphthongs, i.e. /ɔj/, in the same fashion, abbreviating them via compression in PVVA contexts. Thus, there appears to be an association between compressive abbreviation and round nuclei; or, perhaps a prohibition against truncation. Another point supporting this is that among monophthongs examined in this study, only two are associated with compressive abbreviation (at a high level of reliability), both of which are round vowels (see Table 5.2). In contrast to /ɔj/, the two dialect sets handle round-glide diphthongs, i.e. /aw/, in different fashions, with Canadian English allowing truncation of the nucleus while compression is mandated for non-Canadian dialects.

This raises the question of how rounding gestures themselves are coordinated among the various component other (lingual) gestures of round diphthongs. Are they coordinated to one "segment" of the diphthong only, i.e. to only the glide in /aw/ and only the nucleus in /ɔj/? Are they coordinated across the entire gestural complex for both? Or only for one and not the other? Different choices made in this regard may play out as differences in abbreviation patterns across dialects. For example, if a dialect such as Canadian English specifies that rounding gestures originating from glides are not coordinated specifically to the nucleus, then whatever constraints rounding may place on abbrevatory mechanisms will have no effect on what type of abbreviation may occur on the nucleus, thus permitting the onset of /aw/ to be truncated in a similar fashion to /aj/. Or, a dialect such as The West might specify that any vocalic rounding gestures are coordinated to all gestures across a diphthongal gestural complex, and therefore /aw, ɔj/ would be expected to behave in a similar fashion. If these two patterns do in fact occur, we might expect to see variation between them in the presence of coarticulatory lip-rounding of the nucleus (for /aw/) or the glide (for /ɔj/); although I do not have evidence to bear on this question at present, it is certainly something which might be investigated through techniques such as synchronized video recording of lip motion during speech.

A more fundamental question remains, however: Why exactly is lip-rounding associated with compression rather than truncation as an abbrevatory mechanism? Why do rounded vowels not simply truncate when motivated to abbreviate, as with /aj/? Here I can only be even more speculative, but I believe there is some evidence from phonetic studies of lip-rounding which might be brought to bear on this question. Bell-Berti & Harris (1979) examined the production of anticipatory lip-rounding produced during consonants preceding a rounded vowel in English (e.g. in 'sue', as contrasted with 'see') via both acoustic and physical (EMG) measurement. Contra previous findings which had found anticipatory rounding "spreading" up to four consonantal segments in advance of a rounded vowel (Daniloff & Moll 1968), Bell-Berti & Harris determined that anticipatory spreading occurred over a set temporal duration preceding the vowel rather than being a function of the number of segments present, suggesting a coordinated articulatory pattern between the labial and lingual articulations of the vowel, with the labial articulation being activated prior to the lingual articulation.

A following study (Bell-Berti & Harris 1982) examined coarticulation with consonants both preceding and following rounded vowels, and found similar patterns in either direction, such that the vocalic labial gesture was coordinated with the vocalic lingual gesture in a stable temporal pattern, with the labial gesture occurring both prior (anticipatory rounding) and subsequent to (preservatory or carry-over rounding) the lingual gesture. Averaging across the results for all subjects, Bell-Berti & Harris (1982) reported that labial gestures preceded their associated vowel by 160 ms (p. 450). A mean duration for preservatory rounding was not reported, but the data for three of the six subjects (p. 453) indicates a mean carryover duration of 60 ms. These findings suggest that lip-rounding gestures have an overlapping and cohesive effect on associated lingual gestures, and that lip-rounding gestures, in AP terms, may extend durationally *beyond* their associated lingual gestures, in both temporal directions. In terms of PVVA mechanisms, the indication here is that rounding does not so much promote compression as it inhibits truncation via its full overlap of associated non-labial gestures, producing a stabilizing effect on the internal gestural composition of diphthongs.

Another avenue of investigation concerns sequences of lip-rounding. Boyce (1990) looked at the coarticulatory effects of rounding on intervocalic consonants in Turkish and English, using physical (EMG) measurement of articulation. Turkish is a vowel harmony language, such that certain vowel features including rounding are harmonized across syllables within a word. The results indicated that for English, VCV sequences with rounded vowels produced a "trough" or dip in the activity of the musculature producing rounding, as illustrated in Figure 5.19.

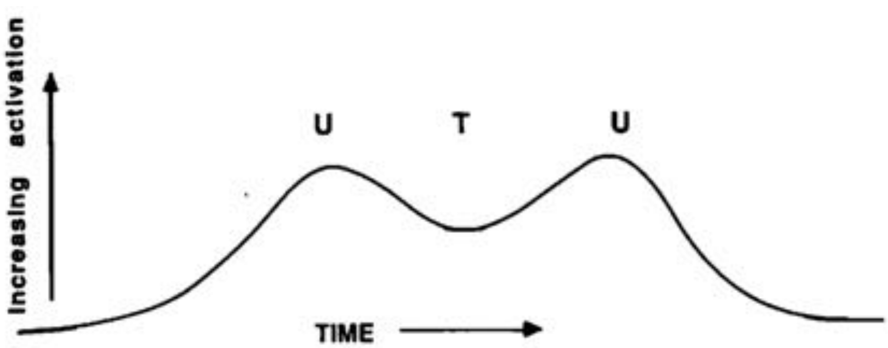


Figure 5.19 Schematized version of "trough" pattern, representing EMG from the orbicularis oris muscle for the utterance /utu/ (Bryce 1990:2584)

Among Turkish speakers, the same VCV sequences involving two rounded vowels "show a consistent plateaulike pattern of movement and a unimodal pattern of EMG activity," (p. 2590). The distinct English and Turkish patterns are compatible with gestural models specifying that English has two discrete labial gestures, each coordinated separately to one of the discrete vocalic lingual gestures, and Turkish having a single labial gesture coordinated with both (and, by extension, all) of the vocalic lingual gestures in a word. These results may not seem obviously relevant to an AP account English lip-rounding gestures, but I think there is an important point to take away from this. Boyce's Turkish data indicate that one-to-many gestural coordination of labial-to-lingual gestures is possible. While English does not exhibit such a coordination pattern across different vowels, coordination of a single labial gesture to multiple lingual gestures *within the same vowel*, i.e. a round diphthong such as /ɔj/, seems a much more reasonable structure to propose for English, given the previous discussion regarding the different patterning of /aw, ɔj/ across dialects.

With the caveat that my account here is much more speculative than prior sections of the PVVA model, I will offer a tentative answer to the question: What do the scores of rounded vowels look like, following the discussion above?

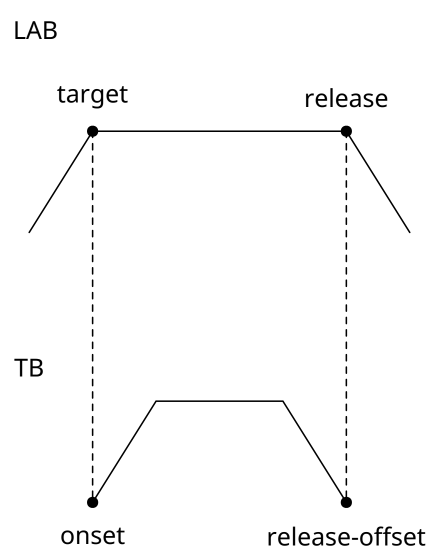


Figure 5.20 Coordination of labial to lingual gesture in round monophthong

To begin, Figure 5.20 provides the basic pattern of coordination between labial and lingual gestures for a simple vowel, a generic monophthong. Following the findings from Bell-Berti & Harris, I coordinate the labial gesture in a manner such that it both precedes and follows the lingual gesture temporally; the specific coordination pattern as illustrated here is perhaps not critical, so long as that basic pattern is achieved.

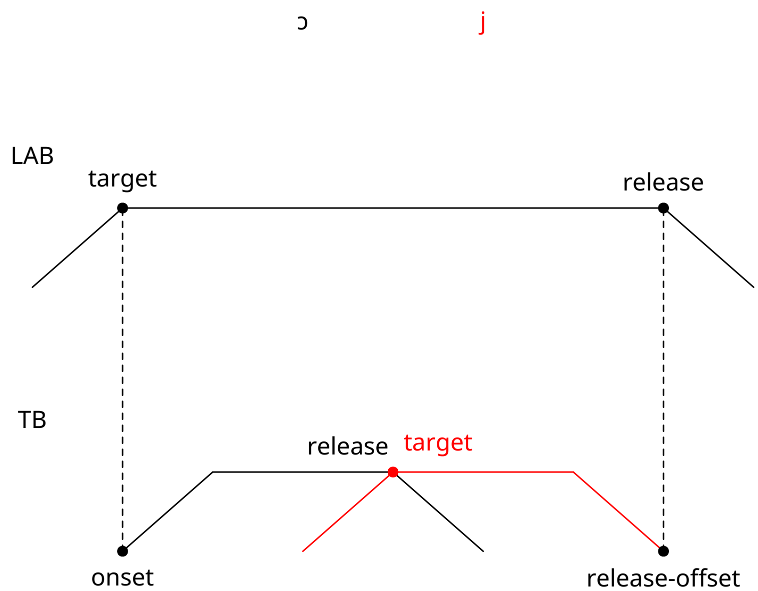


Figure 5.21 Gestural coordination for /ɔj/

From the basic labial pattern, I turn now to the structure of rounded diphthongs, beginning with the vowel /ɔj/, as it has a common pattern across all three dialects. The score in Figure 5.21 combines the basic diphthongal arrangement of the lingual gestures such as shown for /aj/ in Figure 5.6, with the labial coordination pattern shown in Figure 5.20 above.

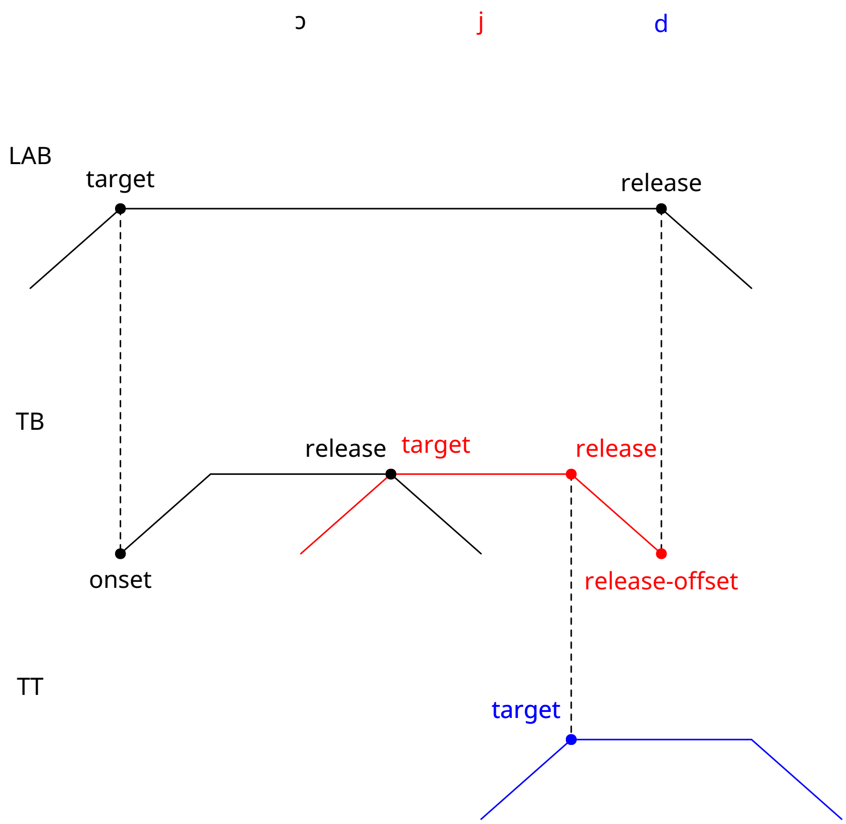


Figure 5.22 Gestural coordination for /ɔjd/

To next illustrate the pre-voiced context, as in /ɔjd/, an additional gesture is added for the following consonant; this also follows a previously illustrated pattern, e.g. in Figure 5.14.

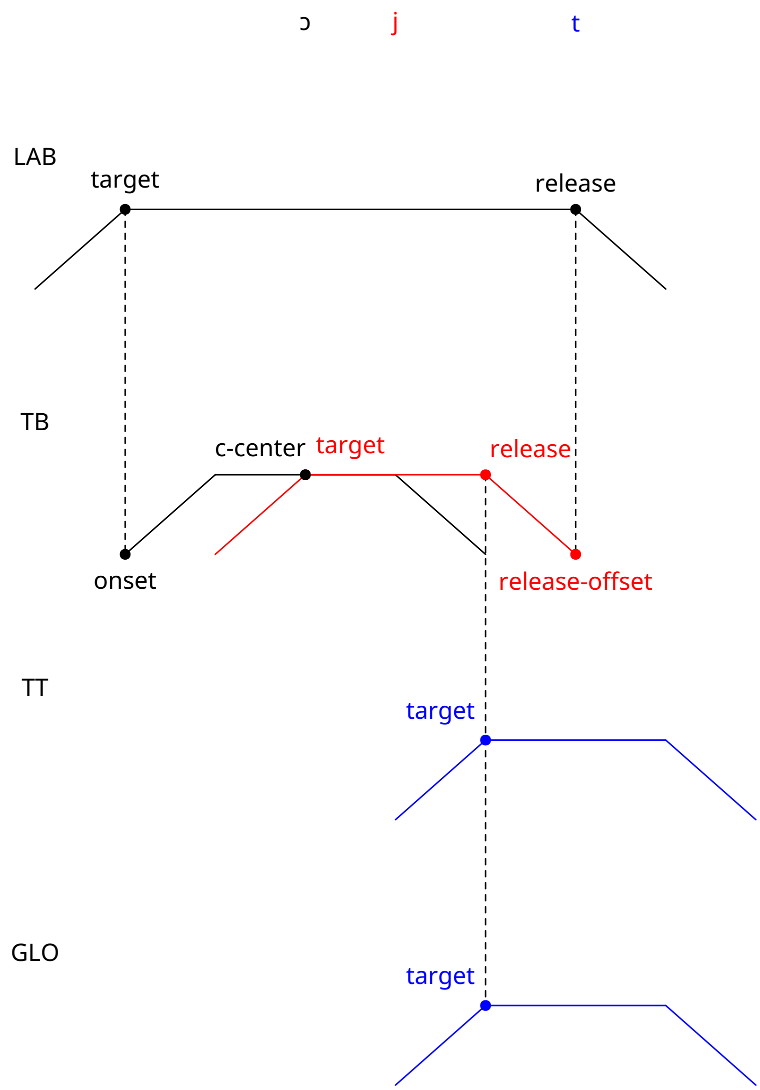


Figure 5.23 Gestural coordination for /ɔjt/; abbreviation via truncation

Assuming for the sake of argument that pre-voiceless /ɔj/ followed the truncation of onset pattern of /aj/, we could now add a glottal gesture and motivate re-alignment of the two vocalic TB gestures, which would produce the configuration in Figure 5.23. This pattern does correctly produce the desired abbreviation, but via the wrong abbrevatory mechanism: truncation of the nucleus. To instead motivate a *compressive* regime for labial vowels, I propose that a π-gesture (on a distinct tier) is associated with all labial gestures.

This is in keeping with the “cohesive” force of lip-rounding, as suggested by Bell-Berti and Harris’ findings. I assume that the stiffness value of this π-gesture is variable, and presumably neutral (i.e. stiffness does not alter) when present in non-pre-voiceless contexts[42]. For reasons which I am not able to fully explicate at present, when a glottal gesture is coordinated to the π-gesture, whether directly or indirectly (I am not certain whether this is an important distinction in the model), the π-gesture *increases* the stiffness factor, thereby producing a *domain of acceleration* which reduces or abbreviates the duration of impacted gestures, as depicted in Figure 5.24.

[42] An alternative explanation might have the π-gesture absent in most contexts, being added only when prompted by the presence of the glottal gesture. I leave evaluation of these two competing accounts (or consideration of other compatible explanations) for further research. The absence of π-gestures in the non-pre-voiceless figures presented here may be taken largely as a simplification for presentation’s sake.

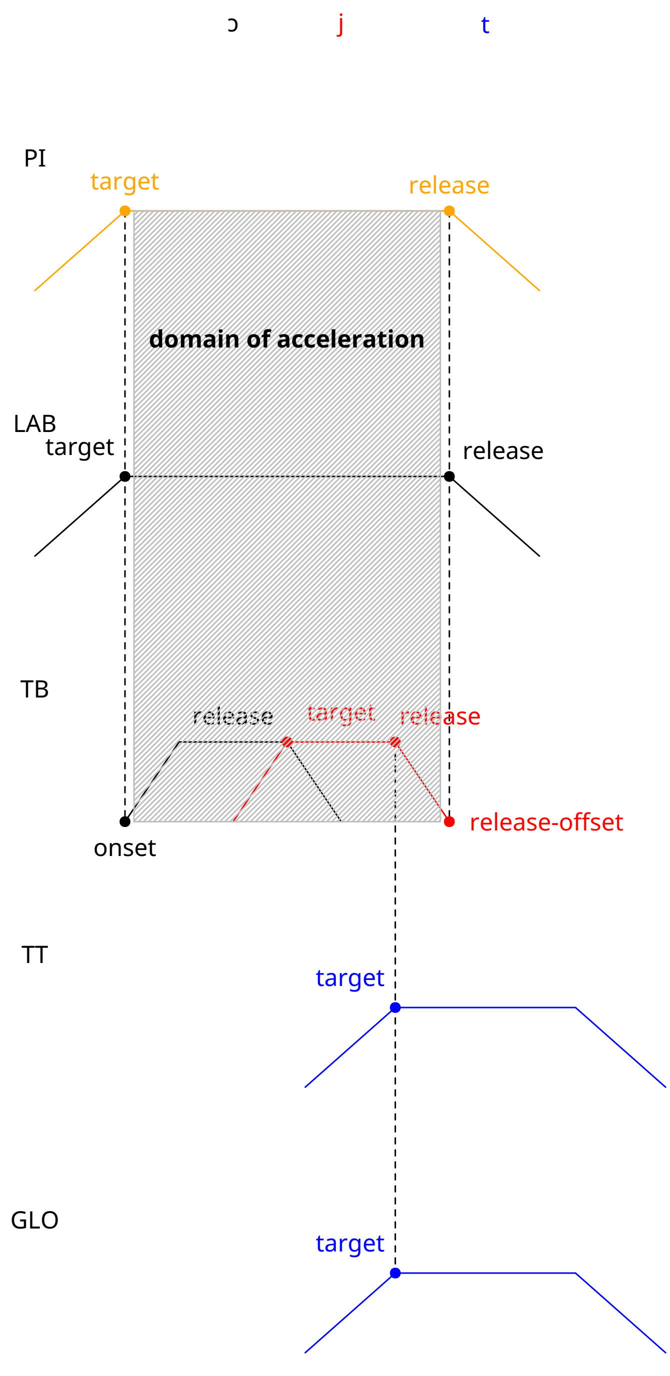


Figure 5.24 Gestural coordination for /ɔjt/; abbreviation via compression

The domain of acceleration produced by the π-gesture in Figure 5.24 results in accelerated cycles for the two TB gestures of the diphthong, thereby producing the desired degree of abbreviation but without altering their internal coordination; the relative proportions of the nucleus and glide remain intact, matching the observations of the data from all three dialects.

For American dialects which have compressive abbreviation of /aw/, its structure will mirror exactly that of /ɔj/ as illustrated in the previous figures. However, for Canadian English, /aw/ instead patterns like /aj/, achieving abbreviation via truncation instead. As suggested earlier, I motivate this by assuming that for round-glide diphthongs in Canada, the labial gesture is not associated with the entire diphthongal complex, but only with the glide's lingual gesture; this pattern is shown in Figure 5.25 for pre-voiced /awd/.

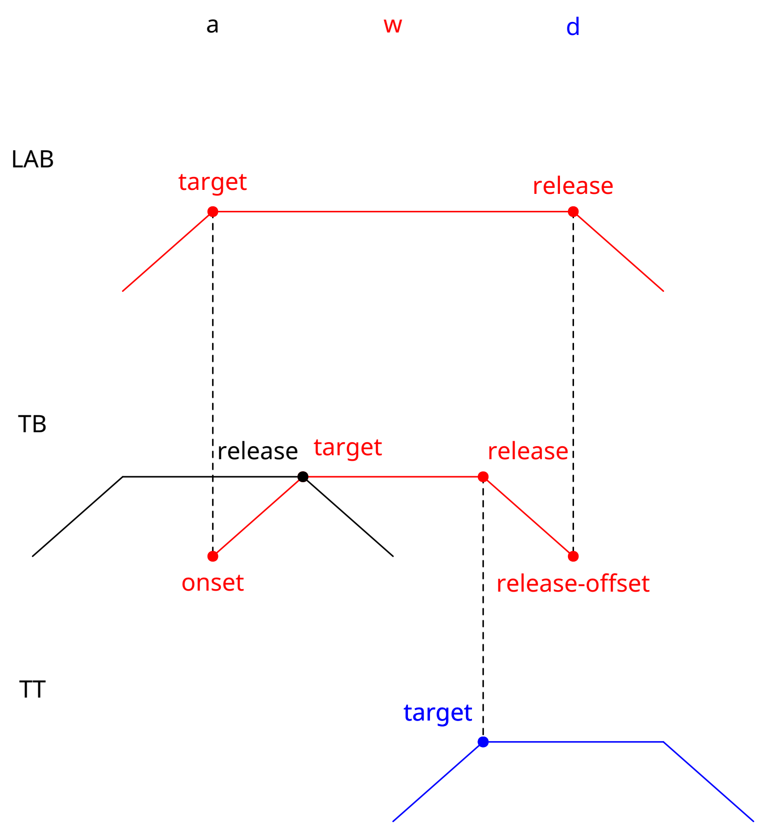


Figure 5.25 Gestural coordination for /awd/; raising dialect (i.e. Canada)

Assuming this basic structure, when a glottal gesture is present as in /awt/, abbreviation can be achieved via realignment of the two lingual gestures, without affecting the coordination between the labial gesture and its sole associated lingual gesture (representing the glide /w/), as shown in Figure 5.26.

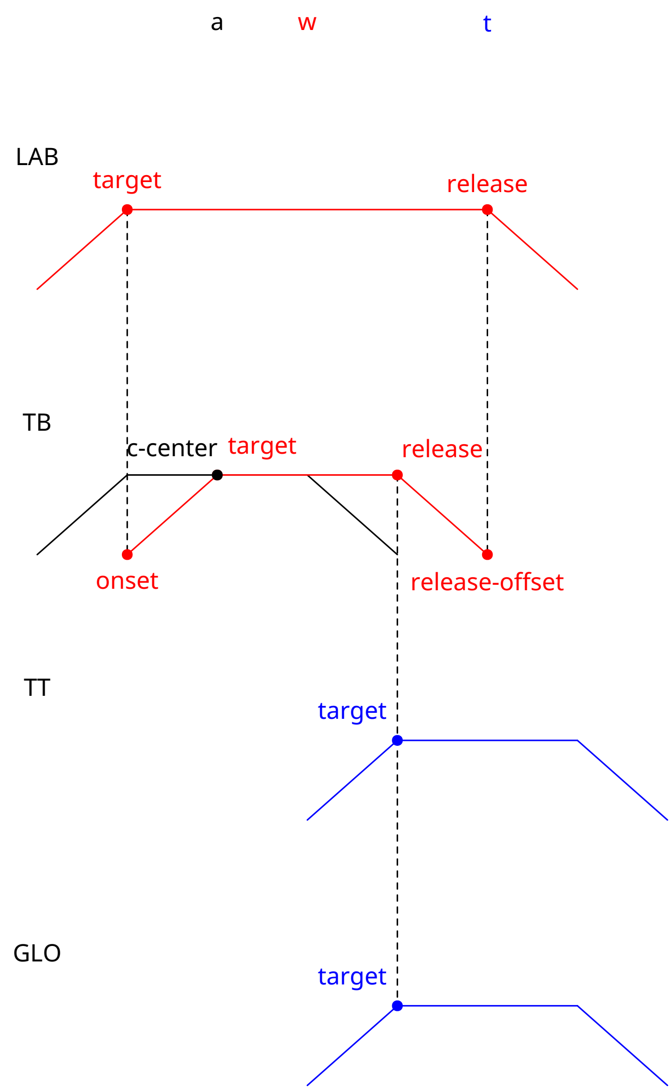


Figure 5.26 Gestural coordination for /awt/; raising dialect (i.e. Canada)

Two important points can be noted following illustration of these proposed structures for /aw/. First, there is a slight but important distinction between Figure 5.22, which applies to /awd/ in non-raising dialects (e.g. The West and The North) and Figure 5.25 for /awd/ in raising dialects (i.e. Canada). In the former the nucleus is fully "covered" by the labial gesture, while in the latter there is a fairly substantial "gap" from the initiation of the lingual gesture until the labial gesture begins. If these structures are accurate, they not only suggest that the nucleus of /aw/ should exhibit some degree of rounding in both raising and non-raising dialects, but also that the temporal extent of that rounding should extend fully across the nucleus for non-raising (i.e. non-Canadian) dialects, while in raising dialects (i.e. Canada) there may be a detectable lack of rounding for some duration during the initial portion of the pre-voiced allophone of /aw/.

In addition to this, just focusing on the two illustrations associated with raising dialects (i.e. Canada), in Figure 5.25 and Figure 5.26, a parallel distinction applies. The slight "gap" between the lingual and labial gestures in pre-voiced /awd/ is not present in pre-voiceless /awt/, and hence the latter should exhibit rounding across a greater proportion of the diphthong. The proposed structures for Canadian /aw/ therefore indicate that there may be a difference between pre-voiced and pre-voiceless /aw/ in terms of the proportion of the nucleus which exhibits rounding, whereas for non-raising (American) dialects the proportion of the nucleus which may exhibit rounding should be relatively similar in either coda voice context. In other words, it may not be that we should necessarily expect to see very much rounding across the board for /aw/ (in any dialect), so much as we might expect to see a *difference* in the rounding of the nucleus, which is correlated with coda voicing, but only in aw-raising dialects, and that such a distinction across coda-voice contexts would be absent in non-aw-raising dialects, which would exhibit the same degree of labiality (whether present or absent) in both contexts.

Given that lip-rounding phenomena are so under-researched in AP, I cannot offer a more comprehensive account of vowel-internal labial gesture coordination at present.

Furthermore, I have not discussed in Chapter 4 any concrete results pertaining to acoustic correlates of lip-rounding, and so do not have any direct evidence to bear on the topic. The discussion in this section does suggest, however, that investigation of diphthongal lip-rounding articulation might be a critical research avenue for developing an AP model of PVVA, such as might be achieved by ultrasound or other physical/visual measurement of lip movement, focusing specifically on the degree and temporal extent of rounding, including dynamic changes over the course of articulation of diphthongs with some degree of rounding, whether associated with the nucleus (/ɔj/) or the glide (/aw/).

5.2.5 IMPLICATIONS OF THE PVVA MODEL

An important question can be raised at this point: Given the PVVA model as described thus far, what types of diphthong patterns *would* be describable and thus expected to occur, and which are more difficult or impossible to handle and thus *would not* be expected? The results from Winnipeg, Denver, and Madison discussed in §4.4.4 identified two of the three possible time-scaling patterns as pertaining to all three diphthongs: either time-normalization, interpreted phonologically here as abbreviation via compression; or, proportional-scaling with right-alignment, interpreted as abbreviation via truncation of the onset, or nucleus. Notably absent is the third pattern, proportional-scaling with *left*-alignment, and its corresponding abbreviation method, truncation of the *offset*, or *glide*.

If truncation of the glide does not occur for diphthongs in the northern dialects, is this merely coincidence, or is there a principled explanation for this pattern within the structure of the model? This can be considered within the context of how the PVVA time-scaling models pattern within the wider vowel inventory. Earlier, I discussed the handling of monophthongs with the specific example of Winnipeg /æ/, which adheres to truncation of the offset (see Figure 5.14 through Figure 5.16). In this case, truncation occurs because of the simplex nature of the monophthongal structure. As there is only one lingual gesture present, coordination of the coda's glottal gesture to an earlier position in the vocalic gesture results in abbreviation of its offset. We may then ask, how do the other monophthongs pattern? Is the left-aligned pattern especially common?

Table 5.4 Cross-dialect comparison of preferred PVVA time-scaling model by vowel, high reliability determinations only (score of 5 or 6)

		Proportionally-scaled model	
City	Time-normalized model	Left-aligned	Right-aligned
Winnipeg	**o**, ɔj	**æ, ʌ**	aj
Denver	aw, ɔj	**ɛ, o**	aj
Madison	**u**, aw, ɔj	—	**ɪ**, aj

Monophthongs in bold

Table 5.2, replicated here as Table 5.4, indicates that there is in fact a strong preference for the truncated, left-aligned pattern among monophthongs (where a high reliability determination can be made), and similarly a strong *dispreference* for the same pattern among diphthongs. As noted earlier, two exceptions exist among the monophthongs, Winnipeg /o/ and Madison /u/. Importantly, these are both round vowels which exhibit a preference for the time-normalization model, interpreted phonologically as abbreviation via compression; they are thus amenable to the labial-gesture motivation for compressive abbreviation as with /ɔj/, discussed in §5.2.4. The only other exception among the monophthongs concerns Madison /ɪ/. There is, however, a strong case to be made that this particular evaluation should have in fact been excluded previously. This is because (see Table 4.8) all of the lax vowels in Madison, including /ɪ/, were found to exhibit *no* *significant correlation* between duration and coda voicing. In other words, although SSANOVA and GAMMs comparisons were run for *all* vowels across all three datasets, the results pertaining to Madison's lax vowels are questionable at best, and more likely should be deemed statistically invalid, and not considered further.

The simplex structure of monophthongs thus indicates that only two abbreviation patterns should pertain under PVVA conditions: *truncation of the offset* when motivated to abbreviate without any present prohibition against truncation, or *compression* when a labial gesture and associated stiffness-increasing π-gesture are also present. I should also reiterate a point made earlier: given their more simplex dynamic structure, most monophthongs are not especially amenable to the formant trajectory comparison methodology conducted in this dissertation, and thus should not be expected to indicate any strong preference for either method. This is indeed the case, as summarized in Table 5.2.

What about the diphthongs? Their more complex structures necessitate consideration of another factor: strategies for dealing with gestural overlap. As discussed in §5.2.2, the northern dialects which exhibit truncation of the nucleus for /aj/ do so for two reasons. First, there appears to be a prohibition against full gestural overlap, such that even under abbreviation regimes (i.e. PVVA) there is still full diphthongality as indicated by the dynamicity of their formant trajectories. Second, I assume that when structures involve partial overlap of diphthong-internal gestures, as motivated by abbrevatory pressures due to the presence of a subsequent glottal gesture, a strategy is adopted which gives precedence to the second gesture, that of the glide (see Figure 5.18).

I do not assume that either of these strategies should be considered a universal approach. So, what if one or both of these two strategies differ across languages, or dialects? Earlier I discussed Marin's (2007a,b) account of Romanian, wherein diphthongal gestures may overlap completely, resulting in a phonetic monophthong. If such a strategy were adopted by an English dialect, what could we expect in terms of phonetic realization? This would depend on the second strategy, how to handle gestural precedence when multiple gestures overlap. Following the northern pattern, a strategy of precedence of the glide might indicate that full overlap should result in a noticeably higher, and fairly or completely monophthongal realization. Conversely, nucleus precedence under full gestural overlap conditions should result in a low monophthong, which is essentially the Romanian pattern. What about cases of only partial gestural overlap, like the northern dialects, but where the *gestural precedence* strategy differs from the northern pattern? That is, cases where non-overlapped portions of both gestures are present, but precedence is given to the nucleus over the glide? Here, we would expect a relatively flat (but not completely monophthongal) diphthong with a long initial steady state, followed by a substantially durationally- and dynamically-reduced off-glide, essentially the mirror image of what we find in northern (Canada, West, North) /aj/ under PVVA conditions. There is a third alternative as well: Marin's account of Romanian diphthongs indicates that partially-overlapped gestures result in monophthongs of *intermediate* quality. Were an English dialect to adopt this strategy, we would expect pre-voiceless allophones to be largely monophthongal but having a significantly raised position relative to their pre-voiced counterpart.

So, the occurrence of different types of diphthongal patterns under PVVA conditions relies to a great extent on the viability of the types of strategies for handling gestural overlap: how much overlap is permitted, and how gestural overlap is handled where it occurs. The dialects investigated for this dissertation reveal that, among English dialects, the following strategies are definitively employed: permit *some* gestural overlap; prohibit *full* gestural overlap; give precedence to the glide over the nucleus where they do overlap. Other potential strategies include: permit full gestural overlap; prohibit *any* gestural overlap; give precedence to the nucleus over the glide where they do overlap. Potentially, these all may occur, in their various logical combinations, in some dialect. Cross-dialect description according to these various strategies, then, is another potentially fruitful area for future investigation. I should emphasize that these strategies, and the patterns which would be associated with them, pertain *only* to PVVA conditions. AP models of non-abbreviation-related diphthongal realizations may be analogous to some of the types of structures resulting from the aforementioned strategies; but, if they are not motivated by some abbrevatory pressure, such as the presence of a glottal gesture (in the case of PVVA), then they cannot be argued to occur as a result of gestural reorganization, and thus do not represent the types of processes discussed here.

5.3 QUALITATIVE DIFFERENCES AS AN OUTCOME OF ABBREVIATION

Throughout this dissertation I have made the argument that CR is directly connected with PVVA, that the acoustic outcomes observed in CR forms are in many respects simply a by-product of the abbrevatory process, despite the fact that PVVA is a pan-English phenomenon, while CR is decidedly not, as traditionally described. The relationship between the two can be summarized as follows:

1. Vowels are abbreviated before voiceless codas (PVVA)

2. There are a number of ways in which vowel abbreviation can occur
3. Other aspects of the phonology place restrictions on the ways abbreviation occurs
4. Canadian English exhibits a particular manner of restriction of abbreviation for the round-off-glide diphthong /aw/ (other diphthongs exhibit patterns which are found in non-Canadian dialects), permitting truncation of the nuclear onset to some degree, which does not occur in other dialects

Within all of this discussion, one important aspect of CR appears to be absent: where is the *raising* in *Canadian Raising*? In other words, what drives the observable (see Table 5.5) qualitative/positional differences between the nuclei of pre-voiceless vs. pre-voiced allophones of /aw/ (CR under my definition) and /aj/ (non-CR, but still similarly affected)? The answer lies in the time-scaling models[43] employed in §4.4, each of which have specific implications for the phonetic outcomes of (articulatory) diphthong production.

The time-normalization model was observed to hold for /ɔj/ in all three dialects examined, and for /aw/ in the two American dialects but not in Canada. In the case of /ɔj/, the outcome of this model is that, despite exhibiting PVVA-induced durational differences between voiced vs. voiceless coda contexts, the phonetic character of the nucleus of /ɔj/ is not affected with respect to position to the extent observed for /aj/ cross-dialectally. This is also largely true for /aw/ in the two American dialects. I believe there are probably a number of factors involved here, some of which these two diphthongs may have in common, and some which are distinct. The obvious common factor (from the present data)

[43] I do not assume here any specific phonological account of the time-scaling models, but do assume that one is possible, such as the AP approach discussed in §5.2.

is the time-scaling method employed in these cases, time-normalization, which can be interpreted in articulatory terms as compression. That is, for these diphthongs in these dialects, the method of achieving PVVA involves compression of the articulatory trajectory, which should result in little or no observable qualitative distinction in acoustic output; and, this is (largely) what is in fact observed.

Concerning /aj/, in all three dialects the proportionally-scaled, right-aligned model was found to account for PVVA differences. In articulatory terms, this implies that abbreviation of /aj/ is achieved by truncating the initial portion of the diphthong, i.e. the nucleus. What is the acoustic outcome of truncation of the initial portion of this articulation? As a low-rising diphthong, the outcome is a *raised nucleus*, exactly what the term *Canadian Raising* (or my proposed alternative, *Northern Raising*) is meant to express.

In the case of /aw/ in Canadian English, its patterns of time-scaling in the face of PVVA, the inferred articulatory adjustment, and the observed acoustic output—a raised nucleus—match those of similarly-time-scaled /aj/. As this is a different pattern than observed for /aw/ in the other dialects, the question must be asked: What is behind this difference? The answer seems likely to be found in a historical change in the phonology of Canadian English with regards to /aw/ (depending on the extent to which Canadian English and the other dialects examined here have similar origins; see Chapter 2). In §5.2.4, I suggest that this may be tied to the round status of the off-glide, which results in a difference in the phonological status of /aw/ in Canadian English tied to its gestural organization, such that PVVA is achieved via a different mechanism than in other dialects, thus aligning /aw/ with /aj/. Have I just argued my way back around to suggesting that CR *is* in fact best defined as the behaviour of /aj, aw/ together in Canadian English? I maintain that this is not the case, because it is the *difference* in the behaviour of /aw/ in Canada, which must have changed at some point in the past (prior to the 20th century; see §2.2), that accounts for the distinctiveness of CR, and not the behaviour of /aj/ which is more widespread cross-dialectally. While /aj, aw/ are united in Canadian English under an abbreviation-via-truncation regime (perceived as "raising"), I have suggested that this is critically tied to the way in which the labial gesture of /aw/ is coordinated to the other gestures, whereas /aj/ lacks labiality entirely. Under this account, it is the status of round-glide gestures, and the difference in gestural coordination between these and round-nucleus gestures (as in /ɔj/) which distinguishes Canadian English from other dialects. The resulting similarities between /aj/ and /aw/ are a byproduct of this more fundamental dialectal distinction, rather than being the main story, as it were.

5.4 On the transcription of diphthongs

In the Introduction, I discussed how the CR diphthongs are typically transcribed, and indicated that I would offer my own views on the best choices in this regard, following the results from my study. Those results, presented in Chapter 4, allow me to now suggest some transcription forms, phonetic and phonological, for the various diphthongs of North American English. In terms of phonetic forms, I proposed in §4.3 a number of transcribed forms based on the results obtained in each dialect, as summarized in Table 5.5.

Table 5.5 Phonetic transcriptions of diphthongs in three dialects of North American English

Diphthong:	/aj/		/aw/		/ɔj/	
Coda:	Voiced	Voiceless	Voiced	Voiceless	Voiced	Voiceless
Canada	[aɪ]	[ʌɪ]	[æʊ]	[ʌʊ]	[oɪ]	[oɪ]
The West	[ɑɪ]	[ʌɪ]	[æo]	[æo̩]	[o̱ɪ]	[oɪ]
The North	[aɪ]	[ʌɪ, ʌe]	[ao]	[ao] ([aö])	[oɪ, oe] ([o̱ɪ, o̱e])	[oɪ, oe]

As any phonetic outcome is subject to variation, I don't believe that the various possible alternate forms presented here need to be debated too strongly; whether pre-voiced /ɔj/ in Madison is best transcribed as [oɪ] or [oe] (or even [o̱ɪ] or [o̱e]) is not worth spilling too much ink over.

Phonological notation is another issue, and one which is probably worth putting a little more focused thought into. There are two components of diphthongal transcription to be discussed here: the nucleus and the off-glide. The off-glide issue was raised in the Introduction. The diphthong notations which I have used throughout this dissertation have used the glide symbols /j, w/, although the proposed phonetic forms in Table 5.1 use vowel symbols instead. I propose that, for phonological purposes, the glides be retained, for two reasons. One, they distinguish the nucleus from the glide, indicating their different phonological statuses. Two, they abstract away from the variety of phonetic forms illustrated in Table 5.1, both within but also between dialects. Assuming that, despite different phonetic outcomes, the three diphthongs represent three distinct phonemes in each dialect (and I see no reason to question that assumption), then a transcription which retains a unified form across dialects of the same language, and fairly close dialects at that, is more useful. Speakers of these three dialects are geographically near to each other, come into contact regularly, are generally exposed to similar types of mass media, and exhibit no issues in terms of mutual intelligibility. Based on such observations, I would argue that, if for no other reason than ease of comparability, these dialects are best described in as similar a fashion as is reasonable. A similar rationale, I believe, lies behind the widely-used *lexical sets* developed by John Wells (1982; see also Wells 2010), which were designed to offer ease of comparability of vowel phonemes across a set of dialects, General American and British Received Pronunciation, whose phonetic realizations are far more distinct than the set of dialects compared in the present study.

Concerning the diphthongal nuclei, there may be some reason to propose an alternative to the common transcriptions which prevail currently, and which I have retained until this point. The traditional forms /aj, aw, ɔj/ suggest that the nuclei of /aj, aw/ are similar in some respect, and that the glides of /aj, ɔj/ are likewise similar. But, there is evidence that, at least among the three dialects under consideration here, that is not entirely the case, and that we might reconsider these notations to reflect observed differences. While /aj, ɔj/ do appear to have reasonably similar final positions in the vast majority of cases and contexts, the nuclei of /aj, aw/ exhibit less similarity. The solution for expressing this dissimilarity is less easily determined. By majority rule among the observations from this limited sample of three dialects, we could retain /aj/ but replace /aw/ with /æw/ instead, indicating that the two diphthongs often do not have similar onset positions, with the caveat that in The North, they have more similar onset positions. Is the solution, then, to utilize /aj, aw/ only in dialects such as The North where they have similar phonetic onsets? This goes against the point I make in the previous paragraph, that a transcription which retains a unified form across relatively similar dialects of a single language is superior to one which does not.

I do not believe this dilemma can be resolved without more data, from a wider range of dialects. I can, however, offer the suggestion of the phonological transcription /æw/ as a way to indicate that it is more distinct from /aj/ than usually understood, with the caveat that more data is needed to more carefully determine the distinctions between the two diphthongs cross-dialectally. And, I would especially emphasize that such data should include an examination of differences between the formant trajectories of these two diphthongs in different-duration contexts, such as I have conducted in this dissertation, to see whether or not they behave distinctly from each other. Although phonological transcriptions such as /aj, æw/ cannot in any real sense capture the complexity of patterning associated with each phoneme, either in their abstract forms or as phonetic output, the choice of notation can and must reflect our understanding of the relationships between phonemes. If research indicates that, by and large, the production patterns of the nuclear portions of these two diphthongs are more or less similar, across a variety of contexts, then retention of /a/ to indicate both of their respective nuclei is warranted; otherwise, distinct representations such as /a, æ/ seems to be a more appropriate choice.

Finally, I will expand the discussion of phonological transcription to the third diphthong of English. Although /ɔj/ is fairly ubiquitous as the phonemic transcription used to indicate for the back, non-low, front-gliding diphthong, the results obtained in this study suggest that /oj/ is just as valid a choice, given that the position of the nucleus is relatively close to monophthong /o/ in each of the three dialects in this study, in both coda contexts examined. Note also that, for dialects such as Canada and The West without /ɔ/ as a distinct monophthongal phoneme, this offers the additional advantage of reducing the number of distinct nuclei within the vowel system. Taking this a step further, I would also note that

/a/ does not occur as a monophthongal phoneme distinct from /æ, ɑ/ in most dialects of English (though for a possible argument against this view, see Boberg 2009 on Canadian *foreign 'a'*). Therefore (and bearing in mind that, as noted earlier, consideration of a broader range of dialects is also necessary), I suggest the following novel set of phonological transcriptions for the three English off-gliding diphthongs, as a potential replacement for the set of currently "standard" forms: /ɑj, æw, oj/. This proposed set offers maximal clarity with regards to cross-diphthongal differences and similarities by differentiating all three diphthongal nuclei, while also minimizing the number of distinct transcribed forms used for vowel nuclei within the overall phonemic inventory.

5.5 OUTSTANDING ISSUES

Here I will summarize a few important outstanding issues pertaining to the topics discussed in this dissertation. One topic which has been raised previously concerns the status of round vowels, and how labiality operates across the duration of diphthongs which have labial components. Another topic of some interest within the wider CR literature which I will briefly comment upon concerns lexical variability in CR application. Finally, I will revisit the issue of flattened diphthongs and other patterns of vowel realization, in the context of the proposed AP model discussed in §5.2.

5.5.1 LABIAL/ROUND VOWELS

One important but potentially controversial aspect of the PVVA model concerns the coordination of labial rounding gestures associated with diphthongs. The model I have proposed implies that labiality or rounding may be associated with portions of diphthongs which have not traditionally been described as round, which is certainly a problem for the model! In particular the following may be implied from the model:

1. The off-glide of /ɔj/ should be round, because the labial gesture associated with /ɔ/ is coordinated across the entire diphthongal structure
2. The nucleus of /aw/ in The West and The North should be round, because the labial gesture associated with /w/ is coordinated across the entire diphthongal structure
3. The nucleus of /aw/ in Canada in pre-voiceless context should exhibit more rounding (perhaps either qualitatively or temporally) than /aw/ in pre-voiced context, because the labial gesture associated with /w/ is coordinated with an earlier position in the nucleus in pre-voiceless context

While I do not have direct data to speak to the truth of any of these implications, when considering these impressionistically I believe that they may be accurate. While I don't think that the off-glide of /ɔj/ necessarily approaches /y/ in its labiality, when comparing /ɔj/ with its decidedly non-round counterpart /aj/, it seems to me that there is certainly a degree of labiality in the former which is lacking in the latter. As /j, y/ are non-contrastive in English, this may be something which is largely unnoticed perceptually. Regarding the third point, while I don't think that the nucleus of /aw/ is necessarily quite /œ/ (or /ɞ, ɔ, ɵ/ etc.), I do feel that impressionistically the rounding associated with /w/ does in fact seem to initiate at an earlier stage in e.g. *lout* vs. *loud* (as a Canadian, I can't speak directly to point #2 above, which concerns strictly American dialects).

Of course, impressionistic views are not enough to substantiate the validity of the implications outlined above. As the status of rounding as a dynamic feature of diphthongal articulations in (North American) English is very much an open question, I think that direct investigation of the rounding status of all three diphthongs across a range of dialects similar to that sampled here (or, at least one Canadian and one American population) would be a very useful avenue for future research in this light.

5.5.2 VARIATION/VARIABILITY IN CANADIAN RAISING

Another issue warranting some comment concerns variation in the application of raising across different lexical items. This has been discussed by Chambers (1973, 1989) inter alia, and often concerns the application of CR to words where factors such as stress, morphological complexity, and syllabification are all involved in differentiating pairs such as *c**i**tation* (no CR) and *c**i**te* (CR). Hall (2005a,b) proposed an account of lexically-conditioned raising in one Ontario community allowing for intermediately-raised forms, arguing that CR is not inherently categorical in nature. I cannot comment too much on Hall's claim of intermediate raising, but I would note that her published data completely exclude the dynamic properties of diphthongs, reporting their formant values only as single positions; regarding duration, Hall comments only that low articulations tend to correlate with longer durations, and states (early in the discussion) that "the role of duration is unclear and is not further examined," (Hall 2005a:6).

In any case, providing an explanation for lexically-conditioned variation is well beyond the scope of this dissertation, as the data which I investigated and attempted to model have strictly concerned monosyllables (and a handful of disyllables with stress on the target syllable). A full AP model for PVVA/CR would certainly need to take into account such factors as stress, syllabic structure, and morphological complexity, but I will not attempt to even begin to outline the necessary requirements for modeling these. Suffice it to say that a fully-conceived model of gestural organization and reorganization would need to potentially deal with different strategies employed in stressed vs. unstressed syllables, and across syllabic and/or morphological boundaries. I will note that a limited investigation of morphological complexity in CR which I carried out in Onosson (2010) as part of an earlier study of Canadian /aj/ found a small but positive correlation between morpheme quantity and vowel duration; but I am unsure whether that relationship has been investigated to any great extent otherwise, especially in the context of CR studies, and so cannot draw any firm conclusions one way or the other on this point.

5.5.3 OTHER PATTERNS: FLAT DIPHTHONGS AND OFF-GLIDING MONOPHTHONGS

As reported in Thomas (2003), a common feature in the American South (though not necessarily restricted to that region; see also Moreton & Thomas 2007) has the three English off-gliding diphthongs "subject to modifications of their glides that might be termed lowering, weakening, or, in some cases, monophthongization" (p. 150); I use the term "flattening" or "flattened" as a general cover for all such processes or descriptions. Thomas describes a variety of diphthong realizations, many of which are *not* amenable to the PVVA model I have described thus far. This is because in such cases, there is no noted relationship to PVVA.

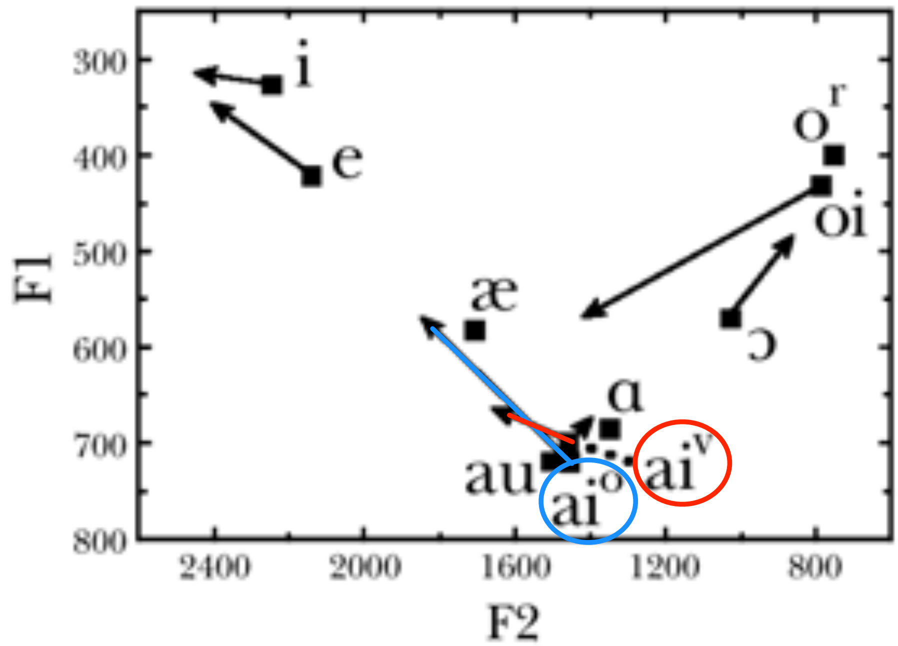


Figure 5.27 Glide Weakening illustrated for an African American Male, Born 1920, from "Springville," Texas (Thomas 2003:153); colour overlays added for clarity

In Figure 5.27, glide weakening patterns for a male, African American Texan born in 1920 are illustrated. Among the /ai/ allophones, superscript V indicates the pre-Voiced context, and a superscript O (for Obstruent) indicates the pre-voiceless context; I have augmented these with coloured circles and trajectory lines, for clarity. Leaving aside the particulars of this speaker's /oi/ diphthong, which exhibits a centralizing trajectory unlike the northern dialects discussed thus far, it is important to note that his /au, oi/ are both non-differentiated by coda voicing context; the same trajectory occurs under both contexts. Although his /au/ is certainly a prime example of a flattened diphthong, in that the nucleus and glide are extremely near each other within the vowel space, it is not one which is subject to the PVVA model I have described previously, which relies on there being gestural *reorganization* in relation to the voice quality of the following coda; here, that simply does not apply.

With respect to the third diphthong, /ai/, the speaker's coda-voice-differentiated allophones follow a similar pattern to the Canadian and more northerly American dialects documented previously, with two major differences: both allophones are notably flattened in both contexts, staying entirely within the lower part of the vowel space; and, his pre-voiceless allophone reaches a substantially higher position at offset; cf. "the more-diphthongal allophone occurs in the short voiceless environment," (Moreton & Thomas 2007:7; see footnote 8). In this case, I would not propose a distinct gestural account, but rather assume that the *positions* indicated by this speaker's off-glide gestures were distinct from those of the northern dialects. Given such distinct positional differences, the AP model described previously, which provides a strategy of precedence for the glide gesture when it overlaps the nucleus, as in pre-voiceless context, might also predict the notably higher offset position of this "flattened" (but not "flat") /ai/ in pre-voiceless vs. pre-voiced position.

Thomas (2003) documents a number of distinct diphthong (and other vowel) realizations across parts of the South, having various patterns beyond those shown in Figure 5.27, including triphthongization and diphthong "breaking". For those vowels which do not pattern allophonically in relation to coda voicing as with /au, oi/ for the speaker discussed above—and these appear to be fairly numerous—there is simply no expected connection to the PVVA model. For those which do exhibit voicing-conditioned allophony, but for which more complex patterns pertain, I would assume that some other gestural features are present which are absent in non-Southern varieties; presumably, a triphthong should contain *three* distinct lingual gestures, for example. Absent direct acoustic data on formant trajectories, vowel durations, etc. from such dialects, I cannot fully account for all such patterns within a framework developed for dialects based on PVVA ratios, diphthong trajectories, and time-scaling models. However, it appears that a somewhat-flattened PVVA diphthong pattern, such as exemplified by /ai/ in Figure 5.27, is not uncommon in the South, and this pattern is certainly amenable to the account I have provided for PVVA under the AP framework. Further data would provide specific direction for determining the inclusion of additional gestural content, for the handling of more complex patterns, indicating that this would be a valuable avenue of research in the future.

I noted earlier that only diphthong realizations with allophonic patterning related to coda voicing, i.e. the PVVA context, should be describable under the proposed PVVA model discussed in §5.2. Leaving this point aside, what can be said about diphthongs which are invariably "flat", such as /au/ in Figure 5.27, or which exhibit other, complex patterns, such as triphthongization. Although the AP model does not necessarily pertain to such structures, aspects of the model which I have described may be applicable to these types of patterns. For example, the generalized diphthong structure which I provide in Figure 5.6 may differ in dialects where e.g. /au/ is barely diphthongal. Whether this entails modeling diphthongs as essentially monophthongs, having only a single gesture, or specifying that the internal structure differs significantly from the "northern" pattern as in Figure 5.6, for example by permitting more complete gestural overlap, I cannot say at present; but certainly, such specifications should be possible.

Similarly, triphthongs are probably best described as complex structures with three distinct gestural components. Here, there are a multitude of possibilities concerning gestural coordination. Having already allowed that the basic internal coordination across two gestures may differ from dialect to dialect, the structural complexities of three coordinated gestures are quite varied. Even allowing just *two* variable landmarks (e.g. c-center, or offset) in a preceding gesture to which a subsequent gesture may be coordinated, there are *four* patterns available for three-gesture structures: offset of G1 with target of G2, and offset of G2 with target of G3; c-center of G1 with target of G2, and offset of G2 with target of G3; offset of G1 with target of G2, and c-center of G2 with target of G3; and, c-center of G1 with target of G2, and c-center of G2 with target of G3. Adding more potential landmarks for coordination, e.g. release-offset (as with labial gestures) multiplies the possibilities again. Whether such patterns are motivated empirically is another question entirely, one which would require a great deal of investigation and which I cannot even begin to answer at present.

Another related topic concerns the off-glides present for some monophthongs. Tense vowels in English such as /i, e, o, u/ are often transcribed as having a notable off-glide within the same relative back-front position, e.g. /ij, ej, ow, uw/; this off-glide is one of the major distinctions between the sets of lax vs. tense vowels, and one of the features that make English vowels rather distinctive compared to a number of other languages. It is reasonable therefore to ask how these off-glides should be represented in a gestural model of vowel production, such as the PVVA model discussed in §5.2. One seemingly obvious possibility would be to incorporate a secondary gesture. This would result in a structure for tense monophthongs resembling or identical to that proposed for diphthongs, such that there would be no phonological distinction between the two. Distinguishing tense monophthongs from diphthongs phonologically might then entail that an additional (third) gesture is required for diphthongs (and a fourth gesture for triphthongs!), which seems at least intuitively to be the wrong route to take. Aside from such intuitions, one of the reasons which I don't believe this is the correct solution is that the single-gesture model of monophthongs which I have proposed nicely predicts that under PVVA, monophthongs should on the whole favour truncation of offset (i.e. proportional-scaling with left-alignment) under PVVA conditions, which turns out to be the case, at least for the present data. Adding a secondary gesture to monophthongal structures would not provide such a prediction, and instead predict that monophthongs should behave in precisely the manner that I currently argue for diphthongs, favouring either truncation of the *onset* or compression when it comes to PVVA. This suggests that another, different explanation should be forthcoming for tense monophthongs which have a notable off-glide.

Regrettably, I am unable to offer a satisfactory gestural account for the distinction between tense vs. lax vowels within the model at present. A more focused analysis of the distinctions between these two categories, which I did not carry out here, might reveal particular differences between these which could elucidate their differential gestural structure in some other way. I have previously noted that the particular method of comparing formant trajectories using different time-scaling models employed in this dissertation may not be especially effective for low-dynamicity vowels, such as monophthongs, even those with off-glides.

Chapter 6 Conclusion

This dissertation has been framed as an attempt to answer the two research questions posed in the Introduction:

1. *What is the contemporary role of duration in the production of Canadian Raising?*
2. *What is the most apt characterization of Canadian Raising?*

Regarding the function of duration in CR, I claim that *Canadian Raising is distinguished specifically by the pattern of durational abbreviation of the diphthong /aw/ in pre-voiceless contexts, which is distinct from its abbreviation patterns in other, non-Canadian dialects, with special attention paid to the dynamic qualities of formant trajectories in abbreviated vs. non-abbreviated allophones*.

This claim is based on the findings of an investigation of duration throughout the vowel system of the dialect of Canadian English spoken in Winnipeg, and two accompanying comparative studies of the English dialects spoken in Denver (The West) and Madison (The North), focusing on the dynamic formant trajectories of diphthongs in those dialects, including the two diphthongs involved in CR: /aj/ and /aw/. That investigation revealed that all three dialects adhere broadly (with some exception in Madison) to a widespread feature of English, pre-voiceless vowel abbreviation, or PVVA, wherein vowels are abbreviated to some degree before voiceless codas. PVVA is not implemented uniformly, but rather varies from vowel to vowel, as well as across dialects, such that the particular arrangement of vowels from least to most affected by PVVA varies between dialects. Among the diphthongs, PVVA patterns differed between the two American dialects on the one hand, which exhibited a (slightly) smaller PVVA effect for /aw/ than for /ɔj/, in comparison to Canada where /aw/ had a substantially larger PVVA effect than /ɔj/. In contrast with /aw/, in all three dialects /aj/ behaved similarly, having a much lower PVVA ratio than either of /aw, ɔj/.

Two methods for statistical comparison of curves, SSANOVA and GAMMs, were employed to identify points of similarity and contrast between PVVA-context diphthongal formant trajectories (as well as throughout the rest of the vowel system). The methodology used in this last part of the study was an attempt at utilizing these tools in a novel manner, to identify which portions of the non-abbreviated trajectories most closely resembled their abbreviated counterparts. The curvilinear comparisons which were conducted under this methodology revealed two broad patterns. The first pattern involves a favourable comparison between per-context trajectories using time-normalization, i.e. where durational differences obtaining from PVVA are effectively ignored when conducting the comparison. The second pattern involves a comparison using proportional time-scaling, i.e. where mean durational differences in each context are taken into account, with abbreviated trajectories being compared to a duration-matched portion of their counterpart non-abbreviated trajectories. The latter comparison has two sub-patterns, given that one of the two compared trajectories contains more datapoints than the other: one sub-pattern has the abbreviated trajectory aligned at the left edge of the non-abbreviated trajectory, and other sub-pattern has it aligned at the right edge. The most favourable comparison model for each vowel, in each dialect, was then determined within each statistical test under the available visual and numeric tools for doing so. These results were compared against each other, and used to provide an estimation of the model which fit the comparison most successfully, ranked on a six-point scale of reliability of the determination. Focusing on the results of this evaluation for the diphthongs, /ɔj/ was consistently found to strongly favour the time-normalization model across all three dialects, whereas /aj/ was found to consistently and strongly favour the proportional-scaling with right-alignment model. For /aw/, two distinct results pertained. In Denver and Madison /aw/, like /ɔj/, strongly favoured time-normalization. However, in Winnipeg /aw/ exhibited a weak preference for proportional-scaling with right-alignment. This latter finding indicates that among the diphthongs, the behaviour of /aw/ in particular stands out as distinguishing Canadian English from geographically proximate and phonetically otherwise fairly similar dialects of the United States.

Regarding the second research question concerning the most apt characterization of Canadian Raising, I propose to describe CR as follows. *Canadian Raising involves the durational abbreviation of the diphthong /aw/ before voiceless codas in a dynamic pattern which is different from other dialects of North American English; the behaviour of the diphthong /aj/, on the other hand, is non-distinctive for Canadian English as compared with a number of other, non-Canadian dialects.* Based upon a review of extant literature (see Table 2.3), it would appear that CR-like patterning of /aw/ is relatively rare as compared with the patterning of /aj/ seen in Canadian English, which is much more widespread in dialects beyond Canada. Therefore, *Canadian Raising is best described as an abbrevatory process which specifically affects /aw/ before voiceless codas, altering its formant trajectories in a particular pattern which is distinct from other dialects*. This pattern of abbreviation of /aw/ in Canadian English is similar to the pattern observed in /aj/

which, being much less dialectally restricted, is therefore not distinctly Canadian. As a common abbrevatory process of vowels before voiceless codas (PVVA) pertains in numerous dialects of English, CR may therefore be understood as one particular implementation, among other possible implementations, of PVVA. The behaviour of /aj/ is a characteristic but non-distinguishing trait of Canadian English, while the similar behaviour of /aw/ in Canadian English is both characteristic, and also distinguishing of Canadian English vs. other dialects. I therefore propose that the term *Canadian Raising* is most usefully and accurately employed in restricted reference to the particular manner in which (only) /aw/ behaves under PVVA in Canadian English[44], and that the behaviour of /aj/ be referred to by another term. Given its occurrence in a number of parts of the central and northern United States (see Table 2.3), *Northern Raising* might be an appropriate and parallel term to describe the abbrevatory patterning of /aj/; alternatively, the simpler and more direct */aj/-raising* is already fairly common, if somewhat prosaic.

Having addressed these major questions, I would like to reflect here upon what has been discussed over the course of this dissertation in relation to common practices in the investigation and description of vowels in sociolinguistics (sociophonetics), dialectology, and other related fields. To echo comments by researchers such as Docherty & Foulkes (1999), Thomas (2002), Jacewicz, Fox & Salmons (2007), I don't think it is an exaggeration to state that by far most the common method of vowel description involves

---

[44] I should emphasize that this does *not* imply that Canadian-like /aw/-raising is in any way restricted to Canada (see Table 2.3); my proposal here is simply to take the pre-existing term *Canadian Raising* and restrict its application in such a way that seems appropriate given the analysis here. While /aw/-raising has been documented in areas such as eastern New England and parts of the U.S. South (Labov 1963; Thomas 2001), it is undoubtedly widespread in Canada, and characteristic of Canadian English (including stereotyping of *Canadian-ness*; see Nycz 2011, 2015). This differentiates /aw/ from /aj/, with which I contrast it here as a means of "deconstructing" Canadian Raising, terminologically.

reporting of mean F1 and F2 values (typically normalized for speaker sex), full stop. Other acoustic qualities, notably including various aspects of vowel duration, including raw duration as well as dynamic spectral changes over that duration, are often not reported at all, or not integrated with the spectral formant data. Only where duration is of special concern, as with the studies on PVVA reported in §2.1, or when investigating topics where it is especially relevant such as diphthongal articulations, does the topic typically arise.

Even where duration does figure into a particular study, it is important to ask, how is it investigated? I will consider here investigations pertaining to diphthong trajectories, as that is obviously the main focus of this dissertation. How many timepoints are captured and considered for examination when comparing formant trajectories? How many are sufficient? The answer to the previous question is, typically, two or three. The answer to the second question is uncertain, although I would suggest that it lies somewhere between three and twenty, the latter being the number of timepoints used in this dissertation. While I don't feel I would have achieved the results I did obtain using only a handful of timepoints, I also don't feel that increasing the number beyond twenty would have substantially changed any of the results; however, the question is worthy of further empirical investigation, certainly.

How are comparisons of (curvilinear) formant trajectories conducted? Here I am considering more the integration of duration and durational differences across diphthongs and phonological contexts, rather than the particular test used for conducting the comparison (e.g. SSANOVA vs. GAMMs). Time-normalization methods appear to be the de facto method of comparison when looking at studies which utilize SSANOVA, for example; in other words, durational differences are immediately discounted. The utilization of such a method should be justified *on its own merits*. I suspect, however, that the choice of time-normalization may often simply be an artefact of the (commonplace) method of data collection, combined with the implementation of SSANOVA (or similar technique) itself. First, considering the implementation of e.g. SSANOVA, it does not appear to be at all common to employ proportional-scaling comparisons such as I have conducted here, much less compare the results of proportional-scaling with those obtained via time-normalization, such as I have carried out here; the same is true of comparisons of variable alignments (left vs. right) within proportional-scaling. Whether or not time-normalization is considered a superior method overall, or simply the best choice for a particular dataset, is something which should be justified as a factor in the methodology of studies which employ it, considering that there are alternatives to doing so. By simply overlooking durational differences, or not considering them in the context of formant dynamicity, it would be easy to not even consider the question in the first place.

Second, consideration of these kinds of methodological concerns needs to be taken into account even during data collection/analysis. Here, I think there is still room for improvement upon the method which I employed in this dissertation. For example, formant values were drawn from my data at twenty timepoints per token *irrespective of durational differences*. For a token with a duration of 100ms, each measurement would therefore be separated by 5ms. But for a longer token with a duration of 200ms, each measurement would be separated by 10ms. To make an analogy with digital sampling, this is equivalent to using a different sampling rate for each token. Is this a reasonable method to employ? I hesitate to admit that I did not even consider that question at the outset of this dissertation, and were I to replicate this study again, I would opt to carry out two sets of measurements—

one at a "variable rate", i.e. using twenty evenly-spaced timepoints, and the other at a "fixed rate", such as every 5 or 10ms. These two sets of measurements could then be used *differentially* when carrying out comparisons using different time-scaling techniques, removing any necessity to scale the data at all. The variable rate measurements could be used for time-normalization comparisons, as is typically done, while the fixed rate measurements could be used for the proportional-scaling measurements, as they would provide precisely the time-matched timepoints desirable for that method. I am not sure whether the results would be substantially different from those achieved under the methodology which I did employ here, where the time-matched "fixed rate" data was reconfigured from the time-normalized "variable rate" data (see §4.4.1), but as the collection of such additional data would not be too onerous—probably involving the change of a single line of code for conducting formant extraction, and a doubling of the numeric spreadsheet data retained for further analysis—it certainly seems like a worthwhile endeavour for future investigations of this type.

While the consideration of acoustic qualities beyond formant measurements, the implementation of multiple timepoints when measuring formants, and the utilization of curvilinear statistical techniques are certainly not novel concepts within the realm of sociophonetics, their combined application in a study such as I have presented in this dissertation appears to be fairly unusual, and may indeed be entirely novel (although I cannot be certain of this). Admittedly, the task of doing so is fairly complicated, and may not even be necessary for many sociophonetic studies which are not of dissertation length. I do hope that my techniques may prove useful to researchers interested in similar questions, or who are dealing with datasets complicated along similar lines as those described here. In particular, I feel that the method I developed for comparing different time-scaling models could be taken up and expanded upon by researchers looking at other aspects of vowel production, examining other languages with complex vowel articulations, and dealing with other types of "messy" data. The potential applications are certainly not limited to sociophonetics, but may be of use to general phoneticians looking at a variety of issues such as: complex consonants with multiple articulations; articulations across consonant clusters; L1 interference patterns in L2 speech; typological differences between languages or dialects; and of course, computational modeling of articulation. I believe there is much room for these kinds of techniques to be refined by such applications, and I hope that they prove of some use to future research so that they can be improved upon and developed, beyond their implementation in this dissertation.

The linguist Fernando Sanchez Miret surveyed a variety of examples of diphthongs, pseudo-diphthongs, diphthongal articulations in monophthongs, "true" vs. "false" diphthongs, etc. From this survey Miret reached the opinion, quoted at the beginning of this dissertation, that "the category 'diphthong' cannot be defined by the presence or absence of some necessary and sufficient conditions of membership. Instead, it is necessary to find a series of features that contribute in different degrees," (Miret 1998:37).

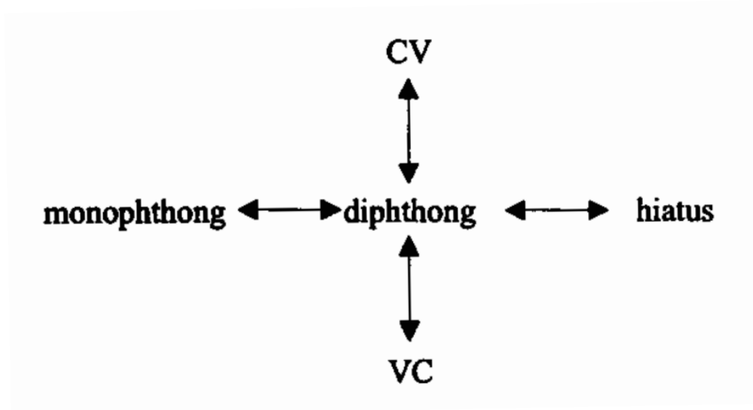


Figure 6.1 Diphthongs at the intersection of phonological spectra in two dimensions (Miret 1998:33)

As shown in Figure 6.1, Miret views diphthongs as occupying a medial space along two dimensions involving spectra of phonological distinctions. When considered along each spectrum, diphthongs are found to exhibit both unitary and dualistic natures. In the horizontal dimension, diphthongs fall between monophthongs and segmental hiatus, i.e. a sequence of two distinct vowels; diphthongs are more dualistic than monophthongs, but more unitary than vocalic hiatus. In the vertical dimension, diphthongs fall between sequences of CV vs. VC, blending characteristics of both vowels and consonants, as well as characteristics of a sequence and a single segment.

Miret ends his discussion of the nature of diphthongs with the following remarks: “In conclusion, diphthongs are complex phenomena that show both unity and duality features, and which can arise in many forms in human language. I think that we come closer to their understanding if we look at their dynamics,” (p. 48). It is within this spirit, *looking closely at the dynamics of diphthongs in order to reach a better understanding of their nature*, that I have carried out the work which comprises this dissertation, and I hope that it has achieved some measure of success in the process.

Bibliography

Abbott, C., Leonard, S. J., & Noel, T. J. (2013). *Colorado: A History of the Centennial State, Fifth Edition*. Boulder: University Press of Colorado.

Ahrend, E. R. (1934). Ontario Speech. *American Speech, 9*(2), 136–139.

Aiello, A. (2010). *A phonetic examination of California.* MA thesis. University of California, Santa Cruz.

Avis, W.S. (1956). Speech differences along the Ontario-U.S. border: III. Pronunciation. *Journal of the Canadian Linguistic Association 2*, 41–59.

Ayearst, M. (1939). A Note on Canadian Speech. *American Speech, 14*(3), 231–233.

Baker, A. (2006). Quantifying Diphthongs. A statistical technique for distinguishing formant contours. Paper presented at *NWAV35*. Columbus OH: Ohio State University.

Beckman, M. (1996). When is a syllable not a syllable? In A. Cutler & T. Otake (Eds.), *Phonological Structure and Language Processing: Cross-Linguistic Studies* (pp. 95–123). Berlin, New York: Mouton de Gruyter.

Beddor, P. S. (2009). A Coarticulatory Path to Sound Change. *Language, 85*(4), 785–821.

Bell-Berti, F., & Harris, K. S. (1979). Anticipatory coarticulation: Some implications from a study of lip rounding. *The Journal of the Acoustical Society of America, 71*(May 1979), 1268–1270.

Bell-Berti, F., & Harris, K. S. (1982). Temporal patterns of coarticulation: Lip rounding. *The Journal of the Acoustical Society of America, 71*(2), 449–454.

Bloomfield, M. W. (1948). Canadian English and Its Relation to Eighteenth Century American Speech. *The Journal of English and Germanic Philology, 47*(1), 59–67.

Boberg, C. (2005). The Canadian shift in Montreal. *Language Variation and Change, 17*, 133–154.

Boberg, C. (2008). Regional Phonetic Differentiation in Standard Canadian English. *Journal of English Linguistics, 36*(2), 129–154.

Boberg, C. (2009). The emergence of a new phoneme: Foreign (a) in Canadian English. *Language Variation and Change*, *21*(3), 355.

Boberg, C. (2010). *The English Language in Canada: Status, History and Comparative Analysis*. Cambridge University Press.

Boersma, P., & Weenink, D. (2017). *Praat: doing phonetics by computer*. URL: http://www.fon.hum.uva.nl/praat/.

Boyens, I. (Ed.) (2007). *The Encyclopedia of Manitoba*. Winnipeg: Great Plains Publications.

Britain, D. (1997). Dialect contact and phonological reallocation: "Canadian Raising" in the English Fens. *Language in Society, 26*(1), 15–46.

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*(1986), 219–252.

Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*(2), 201–251.

Browman, C. P., & Goldstein, L. (1992). Articulatory Phonology: An Overview. *Phonetica, 49*(3–4), 155–180.

Browman, C., & Goldstein, L. (1995). Dynamics and articulatory phonology. In R. F. Port & T. van Gelder (Eds.), *Mind as Motion* (pp. 51–62). Cambridge: MIT Press.

Buja, B. Y. A., Hastie, T., & Tibshirani, R. (1989). Linear Smoothers and Additive Models. *The Annals of Statistics, 17*(2), 453–510.

Byrd, D. (1996). A phase window framework for articulatory timing. *Phonology, 13*(2), 139–169.

Byrd, D. (2003). Frontiers and Challenges in Articulatory Phonology. *ICPhs*, 89–92.

Canadian Association of University Teachers. (2016). *CAUT Guide to Acknowledging Traditional Territory*. https://www.caut.ca/docs/default-source/professional-advice/list---territorial-acknowledgement-by-province.pdf?sfvrsn=12 (accessed September 7, 2017).

Carmichael, K. (2015). "I'm so New Orleans, when I go out of town people ask me if I'm Canadian": Canadian Raising as an innovation in New Orleans English. Oral presentation, *American Dialect Society*. Portland, OR.

Chambers, J. (1973). Canadian Raising. *The Canadian Journal of Linguistics / La Revue Canadienne de Linguistique, 18*(2), 113–135.

Chambers, J. (1975). Canadian Raising. In J. Chambers (Ed.), *Canadian English: Origins and Structures* (pp. 83–100). Toronto: Methuen.

Chambers, J. K. (1981). The Americanization of Canadian Raising. In Masek, C.S., Hendrick, R.A., and Miller, M.F. (Eds.). *Papers from the Parasession on Language and Behaviour.* Chicago Linguistic Society.

Chambers, J. K. (1989). Canadian Raising: Blocking, fronting, etc. *American Speech, 64*(1), 75–88.

Chambers, J. K. (2006). Canadian Raising Retrospect and Prospect. *The Canadian Journal of Linguistics / La Revue Canadienne de Linguistique, 51*(2), 105–118.

Chen, M. (1970). Vowel Length Variation as a Function of the Voicing of the Consonant Environment. *Phonetica, 22*, 129–159.

Chen, Y., & Lin, H. (2011). Analysing tongue shape and movement in vowel production using SS ANOVA in ultrasound imaging. *Proceedings of the 17th International Congress of Phonetic Sciences,* (August), 124–127.

Clarke, S., Elms, F., & Youssef, A. (1995). The third dialect of English: Some Canadian evidence. *Language Variation and Change, 7*, 209–228.

Clopper, C. G., Pisoni, D. B., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America, 118*(3 Pt 1), 1661–1676.

Cooper, C. (1685). *Grammatica Linguae Anglicanae*. London.

Current, R. N. (1976). *The History of Wisconsin Volume II: The Civil War Era, 1848–1873*. Madison: State Historical Society of Wisconsin.

Current, R. N. (1977). *Wisconsin: A Bicentennial History*. New York: W. W. Norton & Company.

Daniloff, R., & Moll, K. (1968). Coarticulation of lip rounding. *Journal of Speech and Hearing Research, 11*, 707–721.

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America, 120*(1), 407–415.

Davis, S., Berkson, K., and Strickler, A. (2016). Incipient /aɪ/-raising in Fort Wayne: The missing dialect? Poster presentation, *NWAV 45*. Vancouver, B.C.

de Jong, K. (1991). An articulatory study of consonant-induced vowel duration changes in English. *Phonetica, 48*, 1–17.

Delattre, P. (1962). Some Factors of Vowel Duration and Their Cross-Linguistic Validity. *The Journal of the Acoustical Society of America, 34*, 1141–1143.

Denes, P. (1955). Effect of Duration on the Perception of Voicing. *The Journal of the Acoustical Society of America, 27*(4), 761–764.

Dollinger, S., & Clarke, S. (2012). On the autonomy and homogeneity of Canadian English. *World Englishes, 31*(4), 449–466.

Docherty, G. J., & Foulkes, P. (1999). Derby and Newcastle: instrumental phonetics and variationist studies. In Foulkes, P. & Dochery, G. J. (Eds.), *Urban voices: Accent studies in the British Isles*, (47–71). London: Routledge.

Durrie, D. S. (1874). *A History of Madison, the Capital of Wisconsin; Including the Four Lake Country to July, 1874*. Madison: Atwood and Culver.

Eckert, P. (2008). Where do ethnolects stop? *International Journal of Bilingualism, 12*(1–2), 25–42.

Elert, C.-C. (1964). *Phonological studies of quantity in Swedish*. Stockholm: Almquist and Wiksell.

Elphinston, J. (1765). *The Principles of the English Language digested: or, English grammar reduced to analogy*. London.

Fintoft, K. (1961). The duration of some Norwegian speech sounds. *Phonetica 7*(1), 19–39.

Fischer-Jørgensen, E. (1964). Sound duration and place of articulation. *Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft, 17*, 175–207.

Fowler, C. A. (1992). Vowel duration and closure duration in voiced and unvoiced stops: there are no contrast effects here. *Journal of Phonetics, 20*, 143–165.

Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech, 36*(2,3), 171–195.

Fox, R. A. & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America 126*(5), 2603–2618.

Fridland, V., Kendall, T., & Farrington, C. (2013). The role of duration in regional U.S. vowel shifts. *Proceedings of Meetings on Acoustics ICA2013, 19*(60296), 1–6.

Fridland, V., Kendall, T., & Farrington, C. (2014). Durational and spectral differences in American English vowels: Dialect variation within and across regions. *The Journal of the Acoustical Society of America, 136*(1), 341–349.

Friesen, G. (1987). *The Canadian Prairies: A History*. University of Toronto Press.

Fruehwald, J. (2008). The Spread of Raising: Opacity, Lexicalization, and Diffusion. *University of Pennsylvania Working Papers in Linguistics, 14*(2), 83–92.

Fruehwald, J. (2013). *The Phonological Influence on Phonetic Change*. PhD dissertation. University of Pennsylvania.

Gafos, A. I. (2002). A Grammar of Gestural Coordination. *Natural Language & Linguistic Theory, 20*(2), 269–337.

Gafos, A. I. (2003). Dynamics: the non-derivational alternative to modeling phonetics-phonology. *Laboratory Phonology, 8*.

Gafos, A., & Goldstein, L. (2012). Articulatory representation and organization. *Oxford Handbook of Laboratory Phonology*, 220–231.

Gandour, J., Weinberg, B., & Rutkowski, D. (1980). Influence of postvocalic consonants on vowel duration in esophageal speech. *Language and Speech, 23*(2), 149–158.

Ghadessy, E. (1986). Some problems of American students in mastering Persian phonology, Unpublished manuscript, University of Texas at Austin.

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. A. Arbib (Ed.), *Action to Language via the Mirror Neuron System* (pp. 215–249). Cambridge: Cambridge University Press.

Greet, W. C. (1931). A Phonographic Expedition to Williamsburg, Virginia. *American Speech, 6*(3), 161–172.

Gregg, R. J. (1973). The Diphthongs əi and aɪ in Scottish, Ulster-Scots and Canadian English. *The Canadian Journal of Linguistics / La Revue Canadienne de Linguistique, 18*(2), 136–145.

Gries, S. T. (2015). The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora, 10*(1), 95–125.

Gu, C. (2002). *Smoothing Spline ANOVA Models*. Springer: New York.

Gu, C. (2014). Smoothing Spline ANOVA Models: R Package gss. *Journal of Statistical Software, 58*(5), 1-25. URL http://www.jstatsoft.org/v58/i05/.

Gu, C., & Wahba, G. (1993). Smoothing Spline ANOVA with Component-Wise Bayesian "Confidence Intervals." *Journal of Computational and Graphical Statistics, 2*(1), 97–117.

Hagiwara, R. (1995). Acoustic realizations of American /r/ as produced by women and men. *UCLA Working Papers in Phonetics 90*.

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *The Journal of the Acoustical Society of America, 102*, 655–658.

Hagiwara, R. (2006). Vowel Production in Winnipeg. *The Canadian Journal of Linguistics / La Revue Canadienne de Linguistique, 51*(2), 127–141.

Hall, E. (2016a). Static and Dynamic Analyses of Canadian Raising in Toronto and Vancouver. Presentation at *New Ways of Analysing Variation 45*, Vancouver.

Hall, E. (2016b). *Static and Dynamic Analyses of Canadian Raising in Toronto and Vancouver*. Unpublished working paper: personal communication.

Hall, K. C. (2005a). *Re-appraising raising: Canadian raising and phonological similarity neighbourhoods*. Unpublished paper: personal communication.

Hall, K. C. (2005b). Defining phonological rules over lexical neighbourhoods: Evidence from Canadian raising. *Proceedings of the 24th West Coast Conference on Formal Linguistics*, 191–199.

Halle, M. (1962). Phonology in Generative Grammar. *Word, 18*(1–3), 54–72.

Hammond, M. (1999). *The Phonology of English: A Prosodic Optimality-Theoretic Approach*. Oxford University Press.

Hastie, T., & Tibshirani, R. (1987). Generalized Additive Models: Some Applications. *Journal of the American Statistical Association, 82*(398), 371–386.

Hastie, T. J.; Tibshirani, R. J. (1990). *Generalized Additive Models*. New York: Chapman and Hall.

Heffner, R.-M. S. (1937). Notes on the Length of Vowels. *American Speech, 12*(2), 128–134.

Heffner, R.-M. S. (1940). A Note on Vowel Length in American Speech. *Language, 16*(1), 33–47.

Heffner, R.-M. S. (1941). Notes on the Length of Vowels (IV). *American Speech, 16*(3), 204–207.

Heffner, R.-M. S. (1942). Notes on the Length of Vowels (V). *American Speech, 17*(1), 42–48.

Heisenberg, W. (1927). Ueber den anschaulichen Inhalt der quantentheoretischen Kinematik and Mechanik. *Zeitschrift für Physik, 43*, 172–198. English translation in Wheeler and Zurek 1983:62–84.

Hewlett, N., Matthews, B., & Scobbie, J. M. (1999). Vowel duration in Scottish English speaking children. In *ICPhS99*, 2157–2160.

Hoffman, M. F., & Walker, J. A. (2010). Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change*, *22*(1), 37–67.

Hogan, J. T., & Rozsypal, A. J. (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *The Journal of the Acoustical Society of America, 67*(5), 1764–1771.

House, A. S. (1961). On Vowel Duration in English. *The Journal of the Acoustical Society of America, 33*(9), 1174–1178.

House, A. S., & Fairbanks, G. (1953). The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels. *The Journal of the Acoustical Society of America, 25*(1), 105–113.

Householder, F. W. (1983). Kyriolexia and Language Change. *Language, 59*(1), 1–17.

Husser, A. (2016, January 10). Ontario's Nipigon River bridge fails, severing Trans-Canada Highway. *CBC News*. Retrieved from: http://www.cbc.ca/news/canada/thunder-bay/nipigon-river-bridge-closed-transcanada-1.3397831 (accessed November 15, 2017).

Jacewicz, E., Fox, R. A., & Salmons, J. (2007). Vowel Duration in Three American English Dialects. *American Speech, 82*(4), 367–385.

Jacewicz, E., & Fox, R. A. (2015). Eliciting Sociophonetic Variation in Vowel Duration. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, 2–6.

Jespersen, O. (1954). *A modern English grammar on historical principles: Part I. Sounds and spellings*. London: George Allen & Unwin Ltd.

Johnson, D. E. (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Linguistics and Language Compass, 3*(1), 359–383.

Joos, M. (1942). A Phonological Dilemma in Canadian English. *Language, 18*(2), 141–144.

Kampstra, P. (2008). Beanplot: A Boxplot Alternative for Visual Comparison of Distributions. *Journal of Statistical Software, Code Snippets 28*(1), 1-9.

Kaye, J. (1990). What ever happened to dialect B? In J. Mascaro & M. Nespor (Eds.), *Grammar in Progress: GLOW Essays for Henk van Riemsdijk* (pp. 259–263). Dordrecht: Foris.

Keating, P. A. (1985). Universal phonetics and the organisation of grammars. In V. A. Fromkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* (pp. 115–132). New York: Academic Press.

Kennedy, R., & Grama, J. (2012). Chain Shifting and Centralization in California Vowels: An Acoustic Analysis. *American Speech, 87*(1), 39–56.

Kilbury, J. (1983). Talking about Phonemics: Centralized Diphthongs in a Chicago-Area Idiolect. In F. B. Agard, G. Kelley, A. Makkai, & V. B. Makkai (Eds.), *Essays in Honor of Charles F. Hockett* (pp. 336–341). Leiden: Brill.

Kingston, J., & Diehl, R. L. (1994). Phonetic Knowledge. *Language, 70*(3), 419–454.

Kinloch, M., & Ismail, F. M. (1993). Canadian Raising: /au/ in Fredericton, New Brunswick. *Linguistica Atlantica, 15*, 105–114.

Klatt, D. H. (1973). Interaction between two factors that influence vowel duration. *The Journal of the Acoustical Society of America, 54*(4), 1102–1104.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America, 59*(5), 1208–1221.

Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: an auditory explanation. *Journal of Phonetics, 16*, 153–169.

Kurath, H. (1977). *A Phonology and Prosody of Modern English*. Ann Arbor: University of Michigan Press.

Kurath, H., & McDavid, R. I. (1965). *The Pronunciation of English in the Atlantic States: Based upon the collections of the Linguistic Atlas of the Eastern United States*. Ann Arbor: The University of Michigan Press.

Labov, W. (1963). The social motivation of a sound change. *Word, 19*(3), 273–309.

Labov, W., Ash, S., & Boberg, C. (2006). *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin, New York: Mouton de Gruyter.

Labov, W., & Baranowski, M. (2006). 50 Msec. *Language Variation and Change, 18*(3), 223–240.

Ladefoged, P. (2006). *A Course in Phonetics (Fifth Edition)*. Thomson Wadsworth.

Ladefoged, P., & Halle, M. (1988). Some Major Features of the International Phonetic Alphabet. *Language, 64*(3), 577–582.

Lehiste, I., & Peterson, G. E. (1959). Linguistic Considerations in the Study of Speech Intelligibility. *The Journal of the Acoustical Society of America, 31*(3), 280–286.

Lehmann, W. P., & Heffner, R.-M. S. (1940). Notes on the Length of Vowels (III). *American Speech, 15*(4), 377–380.

Lehmann, W. P., & Heffner, R.-M. S. (1943). Notes on the Length of Vowels (VI). *American Speech, 18*(3), 208–215.

Lehn, W. (1959) Vowel Contrasts in a Saskatchewan English Dialect. *The Canadian Journal of Linguistics / La Revue Canadienne de Linguistique, 5*(2), 90–98.

Lisker, L. (1975). Is it VOT or a First-Formant Transition Detector? In *Status Report on Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications* (pp. 153–164). New Haven: Haskins Laboratories.

Locke, W. N., & Heffner, R.-M. S. (1940). Notes on the Length of Vowels (II). *American Speech, 15*(1), 74–79.

Lucht, F. (2013). Older Immigrant Languages. In T. Purnell, E. Raimy, & J. Salmons (Eds.) (2013).

Maack, A. (1953). Die Beeinflussung der Sonantendauer durch die Nachbarkonsonanten. *Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft, 7*, 173–178.

Maddieson, I., & Gandour, J. (1976). Vowel length before aspirated consonants. *UCLA working papers in phonetics, 31*, 46-52.

Malécot, A. (1970). The Lenis-Fortis Opposition: Its Physiological Parameters. *The Journal of the Acoustical Society of America, 47*(6), 1588–1592.

Marin, S. (2005). Complex Nuclei in Articulatory Phonology: The Case of Romanian Diphthongs. In Gess, R., Rubin, E. (Eds.), *Selected papers of the 34th LSRL* (pp. 161–177). Amsterdam, Philadelphia: John Benjamins.

Marin, S. (2007a). Vowel to vowel coordination, diphthongs and Articulatory Phonology. PhD dissertation. Yale University.

Marin, S. (2007b). An articulatory modeling of Romanian diphthong alternations. *Proceedings of the XVIth International Congress of Phonetic Sciences, Saarbrücken*, (August), 453–456.

Marin, S. (2014). Romanian diphthongs /ea/ and /oa/: an articulatory comparison with /ja/ - /wa/ and with hiatus sequences. *Revista de Filología Románica, 31*, 83–97.

Marra, G., & Wood, S. (2011). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics, 39*(1), 53–74.

Massaro, D. W., & Cohen, M. M. (1983). Consonant/vowel ratio: an improbable cue in speech. *Perception & Psychophysics, 33*(5), 501–5.

Menclik, C. (2013). *Canadian Raising: Phonetic and Social Factors*. Master's thesis. Universitat Wien.

Metz, C. (1914) *Ein experimentall-phonetishcer Beitrag zur Untersuchung der italienischen Konsonantengemination*. Glückstadt.

Meyer, E. A. (1903). *Englische Lautdauer, eine experimental-phonetische Untersuchung [English length, an experimental-phonetic study]*. Harrassowitz (Leipzig), Uppsala.

Meyer, E. A. & Gombocz, Z. (1909). Zur Phonetik der ungarischen Sprache. *Le Monde oriental 1907-1908*, 122–187.

Milroy, J. (1996). Variation in /ai/ in Northern British English, with Comments on Canadian Raising. *University of Pennsylvania Working Papers in Linguistics, 3*(1), 213–221.

Miret, F. S. (1998). Some reflections on the notion of diphthong. *Papers and Studies in Contrastive Linguistics, 34*, 27–51.

Mitleb, F. (1982). Voicing effect on vowel duration is not an absolute universal. *Journal of the Acoustical Society of America, 71*(S1), S23.

Moisik, S. (2013). *The Epilarynx in Speech*. PhD dissertation. University of Victoria.

Moreton, E. (2004). Realization of the English postvocalic [voice] contrast in F1 and F2. *Journal of Phonetics, 32*(1), 1–33.

Munhall, K., Fowler, C., Hawkins, S., & Saltzman, E. (1992). "Compensatory shortening" in monosyllables of spoken English. *Journal of Phonetics, 20*, 225–239.

Myers, J. (1997). Canadian Raising and the representation of gradient timing relations. *Studies in the Linguistic Sciences, 27*(1), 169–184.

Myers, S. (1987). Vowel Shortening in English. *Natural Language & Linguistic Theory, 5*, 485–518.

Navarro Tomas, T. (1916). Cantidad de las vocales acentuades. *Revista de Filologia Espanola, 3*, 387–408.

Nearey, T. M., & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *The Journal of the Acoustical Society of America, 80*(5), 1297–1308.

Nesbit, R. C. (1973). *Wisconsin: A History, Second Edition*. Madison: The University of Wisconsin Press.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18*(1), 62–85.

Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. *Journal of the American Statistical Association, 83*(404), 1134–1143.

Nycz, J. R. (2011). *Second Dialect Acquisition: Implications for Theories of Phonological Representation*. PhD dissertation. New York University.

Nycz, J. (2015). Second Dialect Acquisition: A Sociophonetic Perspective. *Linguistics and Language Compass, 9*(11), 469–482.

Oliver, E. H. (Ed.). (1915). *The Canadian North-West, Its Early Development and Legislative Records*. Ottawa.

Onosson, S. (2010). *Canadian Raising in Manitoba: Acoustic of effects of articulatory phasing and lexical frequency*. Master's thesis. University of Manitoba.

Onosson, S. (2014). The Prosodic Structure of Canadian Raising. In Teddiman, L. (Ed.) *Proceedings of the 2014 annual conference of the Canadian Linguistic Association.*

Pappas, P. A., & Jeffrey, M. (2014). Raising and shifting in BC English. In A. Barysevich, A. D'Arcy, & D. Heap (Eds.), *Proceedings of Methods XIV: Papers from the Fourteenth International Conference on Methods in Dialectology*, 2011 (pp. 36–47). Frankfurt: Peter Lang.

Paradis, C. (1980). La règle de Canadian Raising et l'analyse en structure syllabique. *The Canadian Journal of Linguistics / La Revue Canadienne de Linguistique, 25*(1), 35–45.

Pater, J. (2014). Canadian raising with language-specific weighted constraints. *Language, 90*(1), 230–240.

Peterson, G. E., & Barney, H.L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America, 24*, 175–184.

Peterson, G. E., & Lehiste, I. (1960). Duration of Syllable Nuclei in English. *The Journal of the Acoustical Society of America, 32*(6), 693-703.

Petty, A. (2013). Immigrant Languages and Education: Wisconsin's German Schools. In T. Purnell, E. Raimy, & J. Salmons (Eds.) (2013).

Picard, M. (1977). Canadian raising: the case against reordering. *The Canadian Journal of Linguistics / La Revue Canadienne de Linguistique, 22*, 144-155.

Port, R. (1976). Influence of tempo on the closure interval cue to the voicing and place of intervocalic stops. The Journal of the Acoustical Society of America, 59(S1), S41–S42.

Port, R. F. (1981). Linguistic timing factors in combination. *Journal of the Acoustical Society of America, 69*, 262-274.

Port, R. F., & Dalby, J. (1982). C/V ratio as a cue for voicing in English. *Perception & Psychophysics, 32*(2), 141–152.

Purnell, T., Salmons, J., & Tepeli, D. (2005). German substrate effects in Wisconsin English: Evidence for final fortition. *American Speech, 80*(2), 135–164.

Purnell, T., Salmons, J., Tepeli, D., & Mercer, J. (2005). Structured Heterogeneity and Change in Laryngeal Phonetics: Upper Midwestern Final Obstruents. *Journal of English Linguistics*, *33*(4), 307–338.

Purnell, T. C. (2009). The Vowel Phonology of Urban Southeastern Wisconsin. *American Speech, 94*(1), 191–217.

Purnell, T., E. Raimy, & J. Salmons (Eds.) (2013). *Wisconsin Talk : Linguistic Diversity in the Badger State*. Madison: University of Wisconsin Press.

Pycha, A., & Dahan, D. (2016). Differences in coda voicing trigger changes in gestural timing: A test case from the American English diphthong /aI/. *Journal of Phonetics, 56*, 15–37.

R Core Team. (2017). *R: A language and environment for statistical computing*. URL: http://www.r-project.org/.

Raphael, L. J. (1972). Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristic of Word-Final Consonants in American English. *The Journal of the Acoustical Society of America, 4*(2), 1296–1303.

Raphael, L. J. (1975). The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English. *Journal of Phonetics, 3*, 25–33.

Roeder, R., & Jarmasz, L.-G. (2010). The Canadian Shift in Toronto. *The Canadian Journal of Linguistics / La Revue Canadienne de Linguistique, 55*(3), 387–404.

Rosen, N., & Skriver, C. (2015). Vowel patterning of Mormons in Southern Alberta, Canada. *Language and Communication, 42*, 104–115.

Rosen, N., Onosson, S. & Li, L. (2015). There's a new ethnolect in town: Vowel patterning of Filipino English in Winnipeg (poster). *New Ways of Analyzing Variation 44*, Toronto.

Rositzke, H. A. (1939). Vowel-Length in General American Speech. *Language, 15*(2), 99–109.

RStudio Team. (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. URL http://www.rstudio.com/

Sadlier-Brown, E. (2012). Homogeneity and autonomy of Canadian Raising. *World Englishes, 31*(4), 534–548.

Sadlier-Brown, E., & Tamminga, M. (2008). The Canadian Shift: Coast to coast. In *Proceedings of the 2008 annual conference of the Canadian Linguistic Association*.

Saltzman, E. (1985). Task dynamic coordination of the speech articulators: A preliminary model. *Status Report on Speech Research SR-84, Haskins Laboratories*, 1-18.

Saltzman, E., & Kelso, J. A. S. (1987). Skilled actions: A task-dynamic approach. *Psychological Review, 94*(1), 84–106.

Saltzman, E. L., & Munhall, K. G. (1989). A Dynamical Approach to Gestural Patterning in Speech Production. *Ecological Psychology, 1*(4), 333–382.

Schilling, N. (2013). *Sociolinguistic Fieldwork*. Cambridge University Press.

Shaiman, S. (2001). Kinematics of compensatory vowel shortening: Intra and interarticulatory timing. *Journal of Phonetics, 29*, 89–107.

Sharf, D. J. (1964). Vowel duration in whispered and in normal speech. *Language and Speech, 32*(1), 89–97.

Shewmake, E. F. (1925). Laws of Pronunciation in Eastern Virginia. *Modern Language Notes, 40*(8), 489–492.

Shewmake, E. F. (1943). Distinctive Virginia Pronunciation. *American Speech, 18*(1), 33–38.

Slis, I. H., & Cohen, A. (1969). On the complex regulating the voiced-voiceless distinction I. *Language and Speech, 12*(2), 80‐102.

Smith, A. E. (1973). *The History of Wisconsin Volume I: From Exploration to Statehood*. Madison: State Historical Society of Wisconsin.

Soli, S. D. (1982). Structure and duration of vowels together specify fricative voicing. *The Journal of the Acoustical Society of America, 72*(2), 366–78.

Sóskuthy, M. (2017). *Generalised Additive Mixed Models For Dynamic Analysis In Linguistics: A Practical Introduction*. Retrieved from https://github.com/soskuthy/gamm_intro

Statistics Canada. (2006). *Aboriginal identity population, by province and territory (2006 Census)*. Retrieved from http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo60a-eng.htm (accessed September 7, 2017).

Statistics Canada. (2011a). *Population and Dwelling Count Highlight Tables, 2011 Census.* Retrieved from http://www12.statcan.ca/census-recensement/2011/dp-pd/hlt-fst/pd-pl/Table-Tableau.cfm? (accessed September 7, 2017).

Statistics Canada. (2011b). *2011 Census of Canada: Topic-based tabulations; Topic-based tabulation: Detailed Mother Tongue.* Retrieved from http://www12.statcan.gc.ca/census-recensement/2011/dp-pd/tbt-tt/Rp-eng.cfm?TABID=2&LANG=E&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=0&GID=908975&GK=0&GRP=1&PID=103251&PRID=10&PTYPE=101955&S=0&SHOWALL=0&SUB=0&Temporal=2011&THEME=90&VID=0&VNAMEE=&VNAMEF (accessed September 20, 2017).

Statistics Canada. (2012). *Table 576-0011 Aboriginal peoples survey, primary Aboriginal language spoken and self-rated ability to speak primary Aboriginal language, by age group and sex, population aged 6 years and over, Canada, provinces and territories. CANSIM (database).* Retrieved from http://www5.statcan.gc.ca/cansim/a26?lang=eng&retrLang=eng&id=5760011 (accessed September 7, 2017).

Statistics Canada. (2016). *Table 051-0056 Population of census metropolitan areas, 2016. CANSIM (database).* Retrieved from http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo05a-eng.htm (accessed September 7, 2017).

Statistics Canada. (2017). *Table A1 Population on April 1st 2017 and population growth, first quarter, 2016 and 2017, Canada, provinces and territories. CANSIM (database).* Retrieved from http://www.statcan.gc.ca/pub/91-002-x/91-002-x2017001-eng.htm (accessed September 7, 2017).

Summers, W. Van. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *The Journal of the Acoustical Society of America, 82*(3), 847–863.

Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change, 24*(2), 135–178.

Tauberer, J., & Evanini, K. (2009). Intrinsic vowel duration and the post-vocalic voicing effect: Some evidence from dialects of North American English. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2211–2214.

The Canadian Encyclopedia. (2017). URL: http://www.thecanadianencyclopedia.ca/en/ Last accessed on August 2, 2017.

Thomas, E. R. (1989). Vowel Changes in Columbus, Ohio. *Journal of English Linguistics, 22*(2), 205–215.

Thomas, E. (1991). The Origin of Canadian Raising in Ontario. *The Canadian Journal of Linguistics / La revue canadienne de linguistique 36*(2), 147–170.

Thomas, E. R. (1995). *Phonetic factors and perceptual reanalyses in sound change.* Doctoral dissertation. University of Texas at Austin.

Thomas, E. R. (2001). *An Acoustic Analysis of Vowel Variation in New World English*. Publication of the American Dialect Society 85. Durham, NC: Duke University Press.

Thomas, E. R. (2002). Instrumental Phonetics. In P. Trudgill, N. Schilling-Estes, & J. K. Chambers (Eds.), *The handbook of language variation and change* (pp. 168–200). Malden MA: Blackwell.

Thomas, E. R. (2003). Secrets Revealed By Southern Vowel Shifting. *American Speech, 78*(2), 150–170.

Tresidder, A. (1941). Notes on Virginia Speech. *American Speech, 16*(2), 112–120.

Tresidder, A. (1943). The Sounds of Virginia Speech. *American Speech, 18*(4), 261–272.

Trudgill, P. (1986). *Dialects in Contact*. New York: Basil Blackwell.

Ubbelohde, C., Benson, M., & Smith, D. A. (2006). *A Colorado History, Ninth Edition*. Boulder: Pruett Publishing Company.

Umeda, N. (1975). Vowel duration in American English. *The Journal of the Acoustical Society of America, 58*(2), 434–445.

University of Victoria. (2017). *Territory acknowledgment*. URL: https://www.uvic.ca/services/indigenous/facultystaff/territory-acknowledgment/index.php. Last accessed: October 10, 2017.

U.S. Census Bureau. (2015). *S1601 Language spoken at home; 2011-2015 American Community Survey 5-Year Estimates*. Retrieved from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=CF. Last accessed: September 21, 2017.

van Rij, J., Wieling, M., Baayen, R. and van Rijn, H. (2016). "itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs." R package version 2.2.

Wardrip–Fruin, C. (1982). On the status of temporal cues to phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants. *The Journal of the Acoustical Society of America, 71*(1), 187.

Wassink, A. (2013). SSANOVA-R-Code-NWAVdemo-rev. R script. Retrieved from: http://faculty.washington.edu/wassink/NWAV2013/NWAV-demo-SSANOVA-RCode-new.txt. Last accessed: February 21, 2017.

Wells, J. (1982). *Accents of English, 3 volumes*. Cambridge: Cambridge University Press.

Wells, J. (2010). lexical sets. Retrieved from: http://phonetic-blog.blogspot.ca/2010/02/lexical-sets.html. Last accessed: January 17, 2018.

Wheeler, J.A. & Zurek, W.H. (eds.). (1983). *Quantum Theory and Measurement*. Princeton University Press.

Whitehead, R. L., & Jones, K. O. (1976). The effect of vowel environment on duration of consonants produced by normal-hearing, hearing-impaired and deaf adult speakers. *Journal of the Acoustical Society of America, 60*(2), 513–515.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the Americal Statistical Association, 99*, 673–686.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Boca Raton: CRC Press.

Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B) 73*(1), 3-36.

Wood, S. N. (2013a). On p-values for smooth components of an extended generalized additive model. *Biometrika, 100*(1), 221–228.

Wood, S. N. (2013b). A simple test for random effects in regression models. *Biometrika, 100*(4), 1005–1010.

Xu, Y. (2007-2015). *FormantPro.praat v. 1.4* [Praat script]. Retrieved from: http://www.phon.ucl.ac.uk/home/yi/FormantPro/. Last accessed: December 2, 2016.

Yuan, J. & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. In *Proceedings of Acoustics '08*.

Zimmerman, S. A. & Sapon, S. M. (1958). Note on Vowel Duration Seen Cross-Linguistically. *The Journal of the Acoustical Society of America, 30*, 152–153.

Appendix A Elicitation wordlist

This appendix presents the wordlist that was used for data collection in as described in §3.1, in two formats. First, as an alphabetical list:

> *adroit, avoid, boil, bout, bowed, bribe, choice, coif, coin, couch, cow, cowl, coy, drive, fine, foist, fount, gouge, had, hade, hah, hat, hate, haw, he, head, heat, heed, height, heist, het, hey, hid, hide, high, hit, hod, hoe, hoed, hooed, hook, hoot, hot, hote, hour, houst, how, hoy, hoyed, hoyt, hud, huh, hut, kine, lies, like, lithe, loud, mouse, mouth, mouthe, oblige, pine, pint, point, poise, pout, pyre, rice, ride, rife, right, rile, ripe, rouse, sky, time, tine, tout, town, toys, who, writhe*

Secondly, as a table showing the various syllabic contexts in which vowels were included. For the monophthongs, these were limited to initial /h/ and final /t, d/ or an open syllable. For diphthongs, contexts were much more varied, drawing from Hammond's (1999) survey of all existing combinations in which the diphthongs are found in genuine words. An 'X' in the Frame column in the table indicates that any segment was permitted in that position. An empty cell indicates that no existing word fills the frame, it is deemed impossible for such a form to exist in English, and/or it was difficult to provide a nonce orthographic form which would be likely to elicit the desired pronunciation (also as noted, in the case of monophthongs, only the first three frames were utilized, so a large portion of the table is empty).

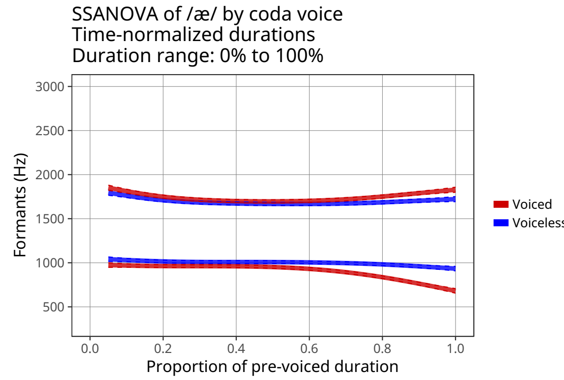
Frame	aj	aw	ɔj	i	ɪ	e	ɛ	æ	ɑ	ʌ	o	ʊ	u
h __ t	height	bout	hoyt	heat	hit	hate	het	hat	hot	hut	hote	hook	hoot
h __ d	hide	bowed	hoyed	heed	hid	hade	head	had	hod	hud	hoed	hood[45]	hooed
h __ #	high	how	hoy	he		hey		hah	haw	huh	hoe		who
p __ X	pine	pout	poise										
t __ X	tine	tout	toys										
k __ X	kine		coin										
X __ #	sky	cow	coy										
X __ p	ripe												
X __ t	right	bout	adroit										
X __ k	like												
X __ b	bribe												
X __ d	ride	loud	avoid										
X __ f	rife		coif										
X __ θ	lithe	mouth											
X __ s	rice	mouse	choice										
X __ v	drive												
X __ ð	writhe	mouthe											
X __ z	lies	rouse	poise										
X __ m	time												
X __ n	rine	town	coin										
X __ l	rile	cowl	boil										
X __ ɹ	pyre	hour											
X __ tʃ		couch											
X __ dʒ	oblige	gouge											
X __ st	heist	houst	foist										
X __ nt	pint	fount	point										

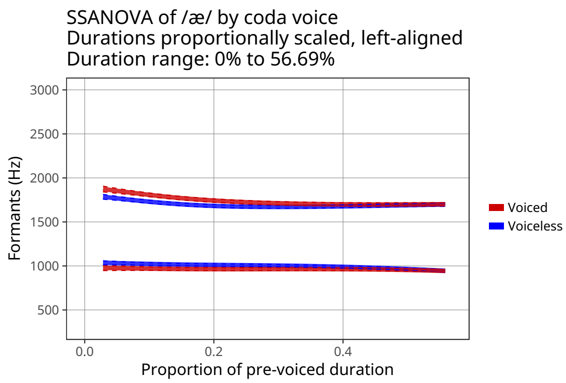
[45] This word was erroneously omitted from the elicitation materials for the Winnipeg study. Although this was corrected for the elicitation materials used in the subsequent studies in Denver and Madison, the vowel /ʊ/ was omitted from the analysis presented in this dissertation so as to permit comparison between equivalent datasets across all three studies.

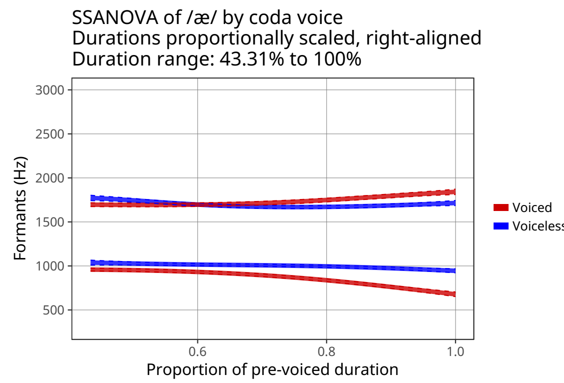
Appendix B SSANOVA comparisons, Winnipeg

This appendix contains the full set of SSANOVA comparisons for each durational scaling /alignment model (see §4.4.2) across all vowels in the Winnipeg dataset. The time-normalized comparison model is listed at the top for each vowel, followed by the proportionally-scaled duration comparison models below, with left- and right-alignments, respectively.

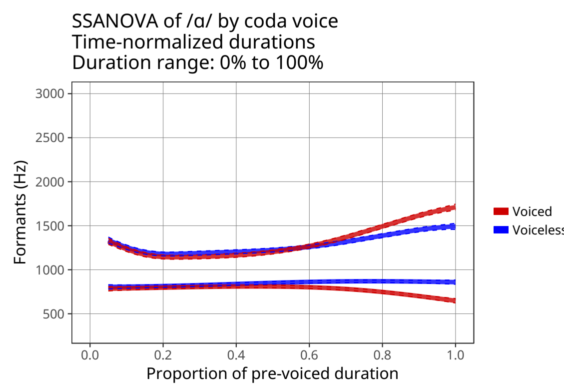
æ

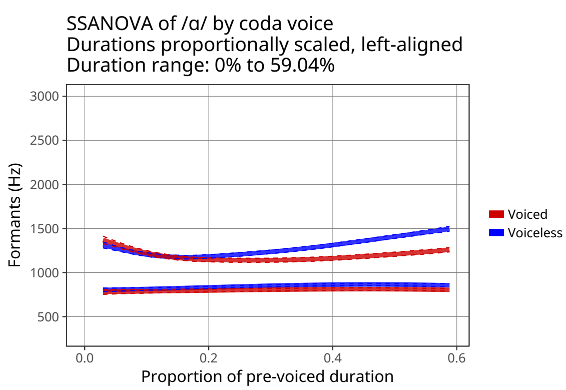

SSANOVA of /æ/ by coda voice
Time-normalized durations
Duration range: 0% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

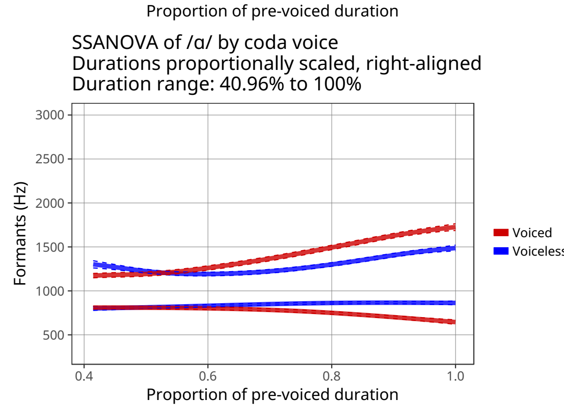

SSANOVA of /æ/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 56.69%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless


SSANOVA of /æ/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 43.31% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

ɑ

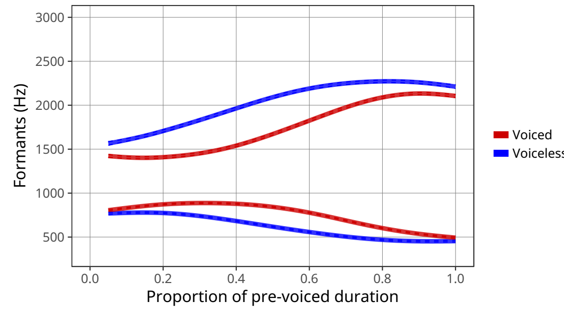

SSANOVA of /ɑ/ by coda voice
Time-normalized durations
Duration range: 0% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless


SSANOVA of /ɑ/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 59.04%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

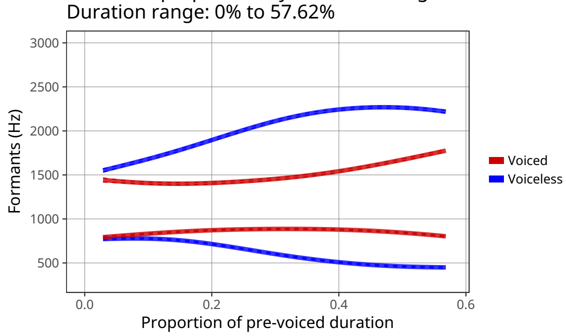

SSANOVA of /ɑ/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 40.96% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

aj

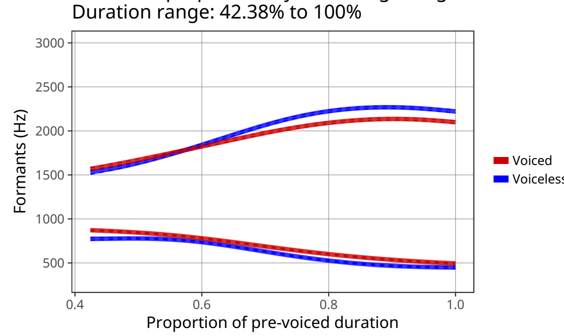
SSANOVA of /aj/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



SSANOVA of /aj/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 57.62%

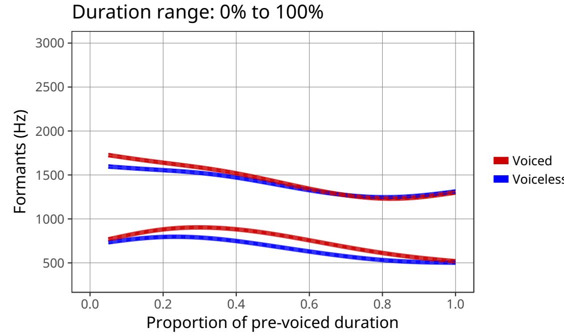


SSANOVA of /aj/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 42.38% to 100%

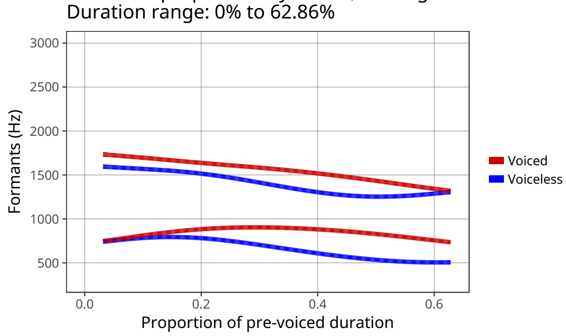


aw

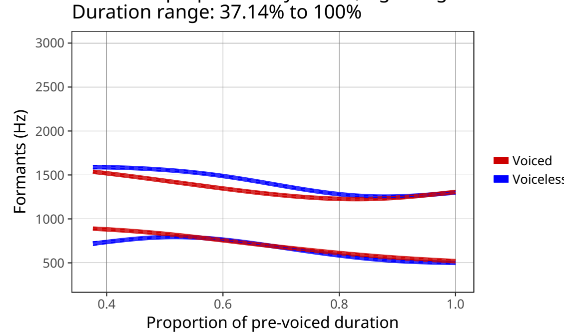
SSANOVA of /aw/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



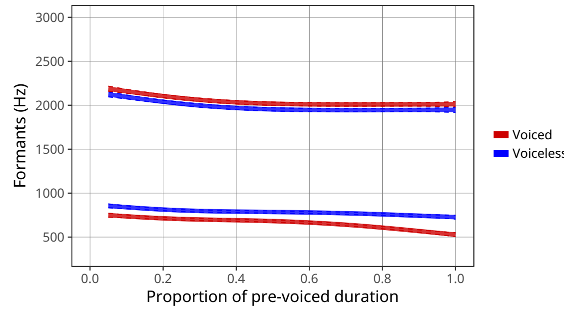
SSANOVA of /aw/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 62.86%



SSANOVA of /aw/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 37.14% to 100%

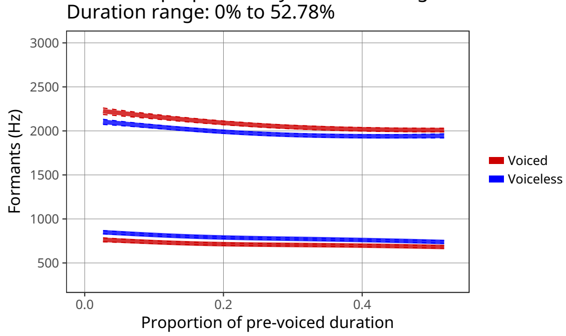


SSANOVA of /ɛ/ by coda voice
Time-normalized durations
Duration range: 0% to 100%

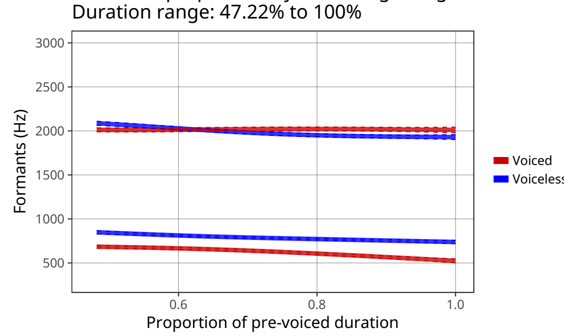


ɛ

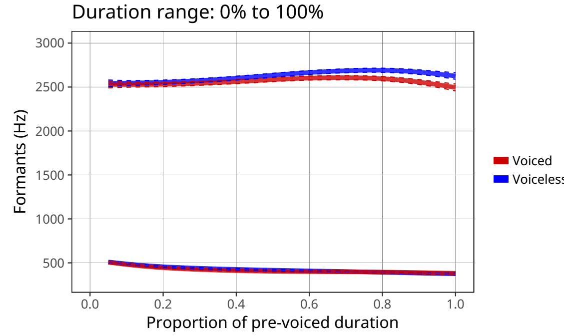
SSANOVA of /ɛ/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 52.78%



SSANOVA of /ɛ/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 47.22% to 100%

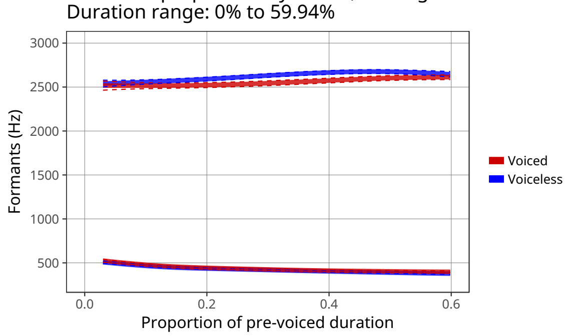


SSANOVA of /e/ by coda voice
Time-normalized durations
Duration range: 0% to 100%

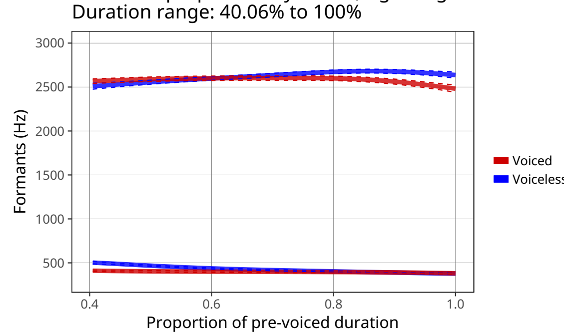


e

SSANOVA of /e/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 59.94%

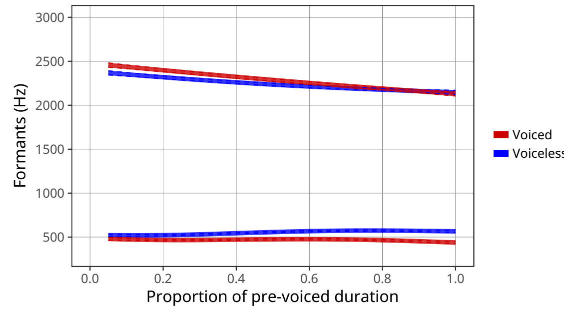


SSANOVA of /e/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 40.06% to 100%

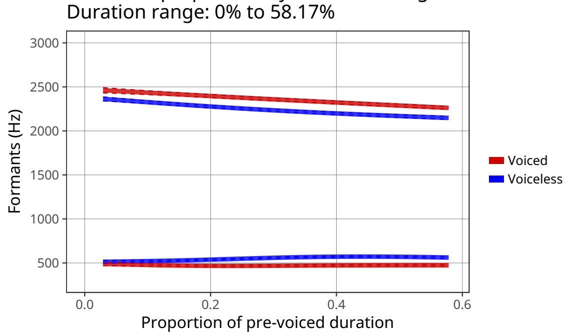


ɪ

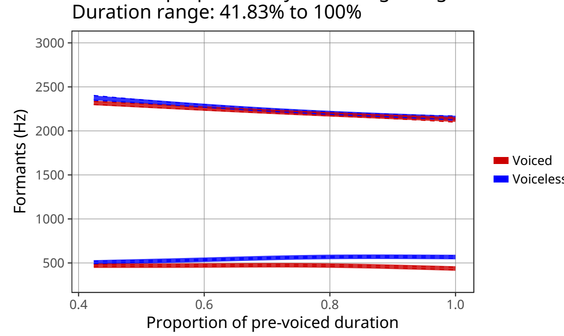
SSANOVA of /ɪ/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



SSANOVA of /ɪ/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 58.17%

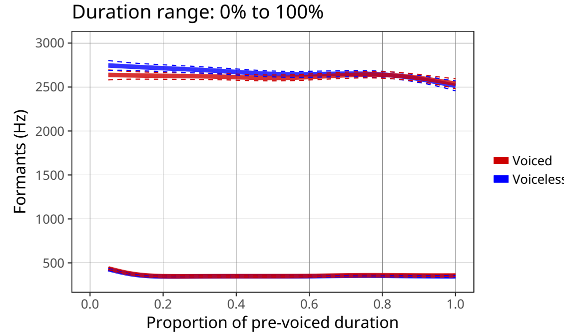


SSANOVA of /ɪ/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 41.83% to 100%

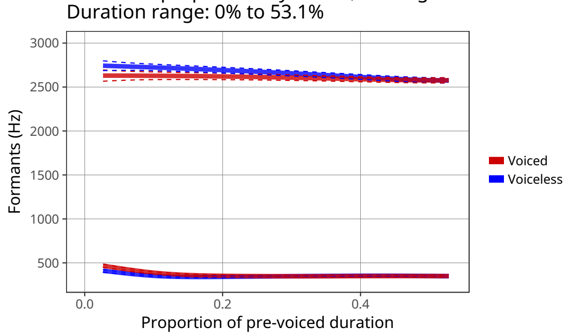


i

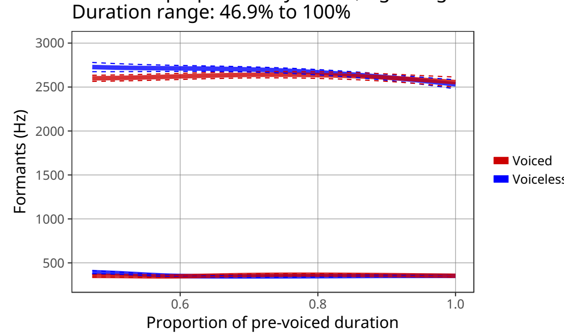
SSANOVA of /i/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



SSANOVA of /i/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 53.1%

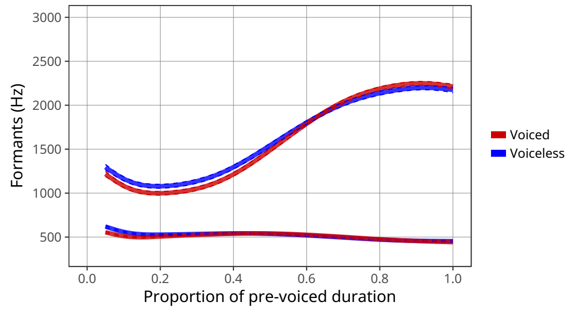


SSANOVA of /i/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 46.9% to 100%

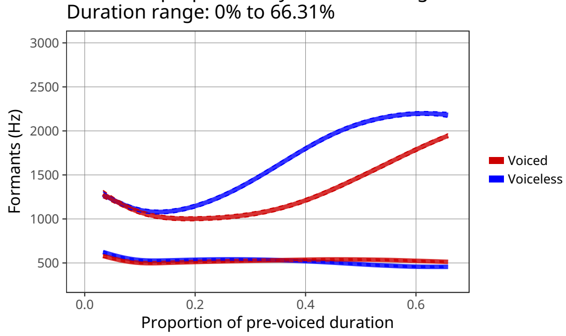


ɔj

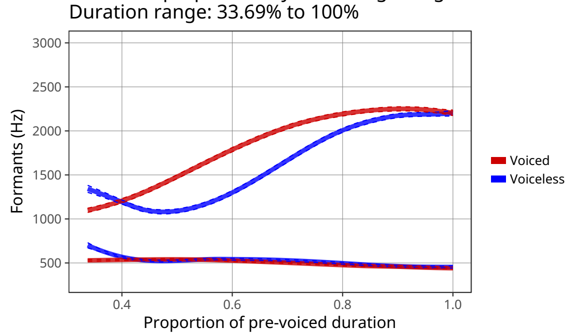
SSANOVA of /ɔj/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



SSANOVA of /ɔj/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 66.31%

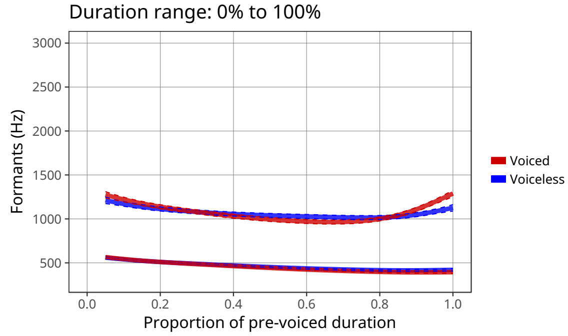


SSANOVA of /ɔj/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 33.69% to 100%

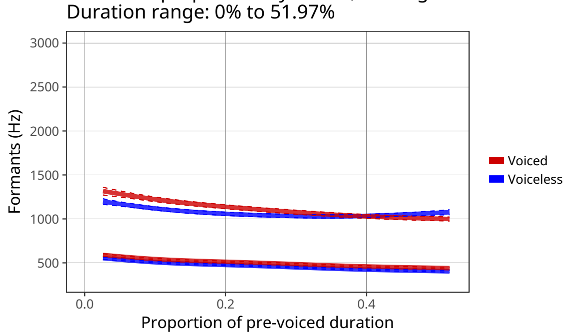


o

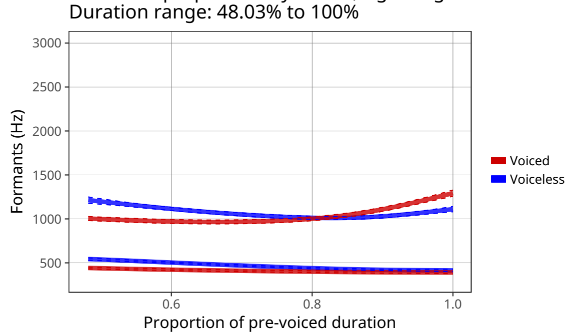
SSANOVA of /o/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



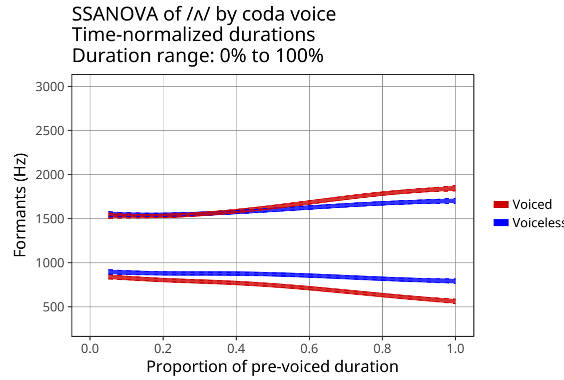
SSANOVA of /o/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 51.97%

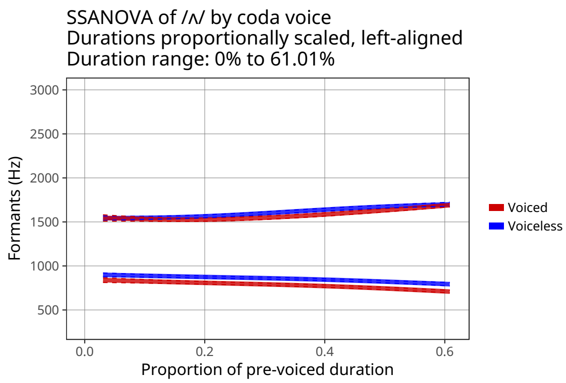


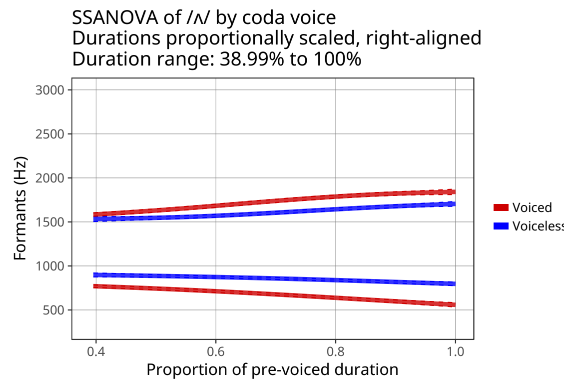
SSANOVA of /o/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 48.03% to 100%



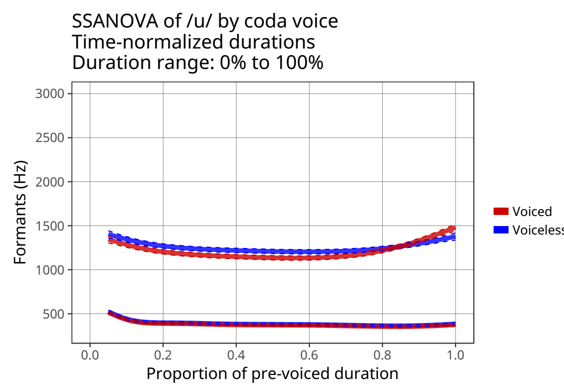
ʌ

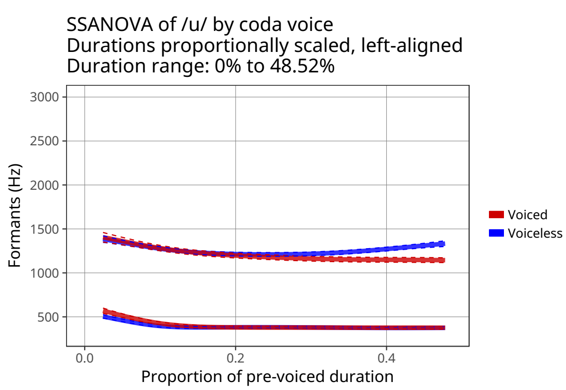

SSANOVA of /ʌ/ by coda voice
Time-normalized durations
Duration range: 0% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

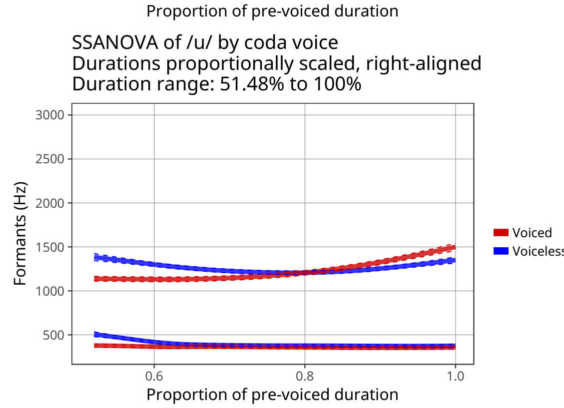

SSANOVA of /ʌ/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 61.01%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless


SSANOVA of /ʌ/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 38.99% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

u


SSANOVA of /u/ by coda voice
Time-normalized durations
Duration range: 0% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

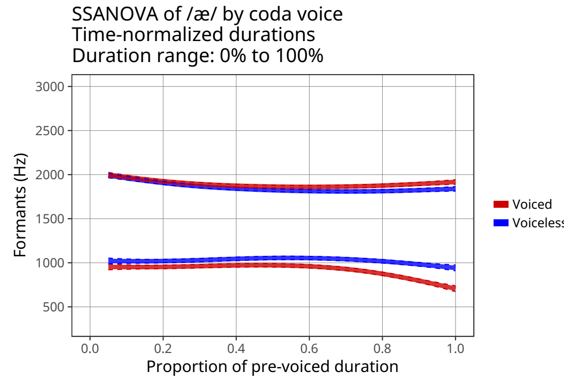

SSANOVA of /u/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 48.52%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

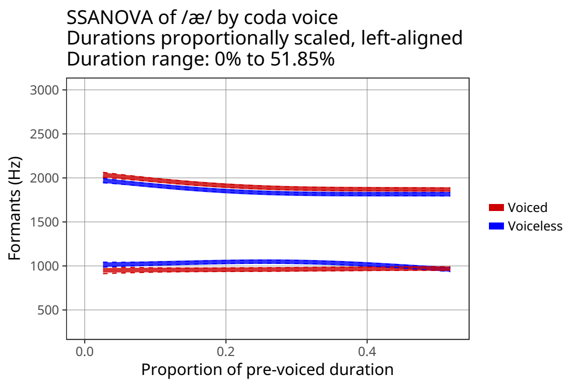

SSANOVA of /u/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 51.48% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

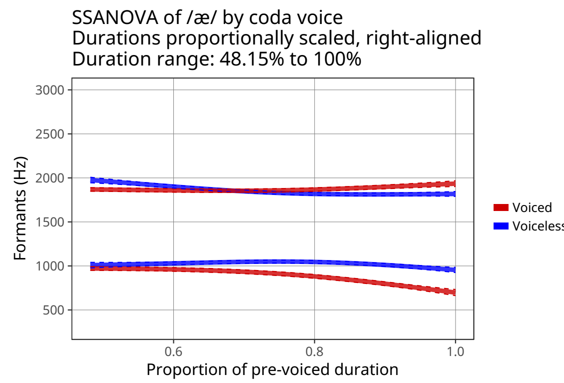
Appendix C SSANOVA comparisons, Denver

This appendix contains the full set of SSANOVA comparisons for each durational scaling /alignment model (see §4.4.2) across all vowels in the Denver dataset. The time-normalized comparison model is listed at the top for each vowel, followed by the proportionally-scaled duration comparison models below, with left- and right-alignments, respectively.

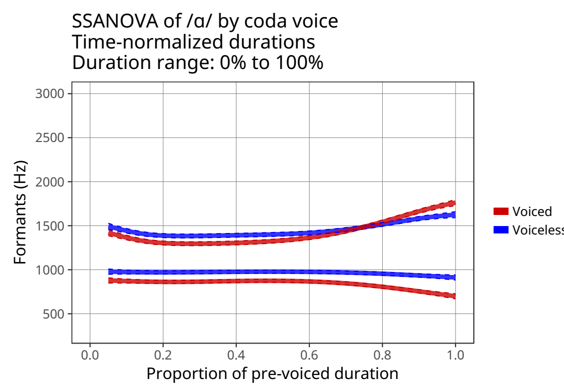
æ

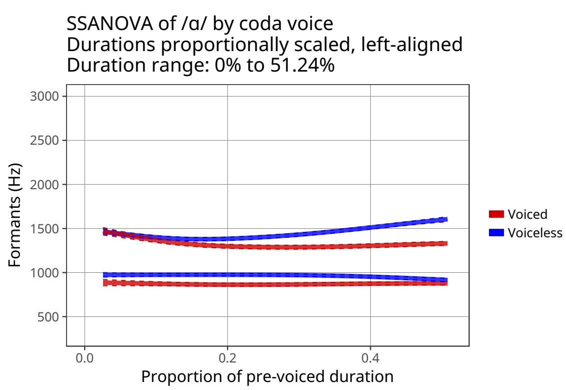

SSANOVA of /æ/ by coda voice
Time-normalized durations
Duration range: 0% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

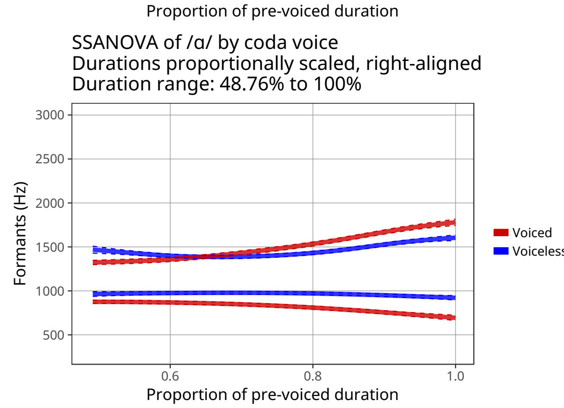

SSANOVA of /æ/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 51.85%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless


SSANOVA of /æ/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 48.15% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

ɑ

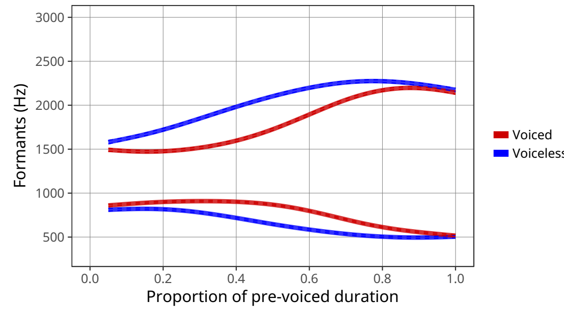

SSANOVA of /ɑ/ by coda voice
Time-normalized durations
Duration range: 0% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless


SSANOVA of /ɑ/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 51.24%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

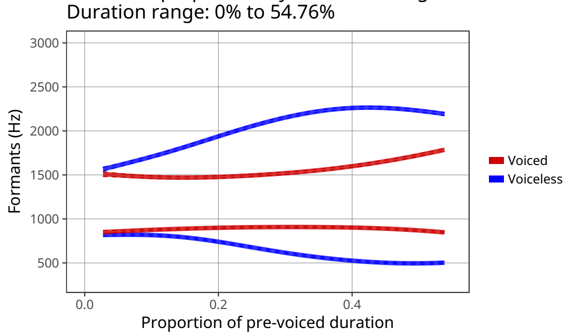

SSANOVA of /ɑ/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 48.76% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

aj

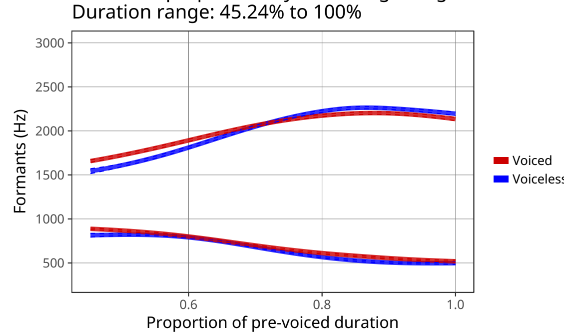
SSANOVA of /aj/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



SSANOVA of /aj/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 54.76%

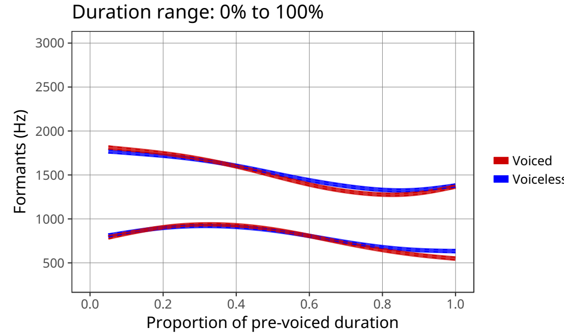


SSANOVA of /aj/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 45.24% to 100%

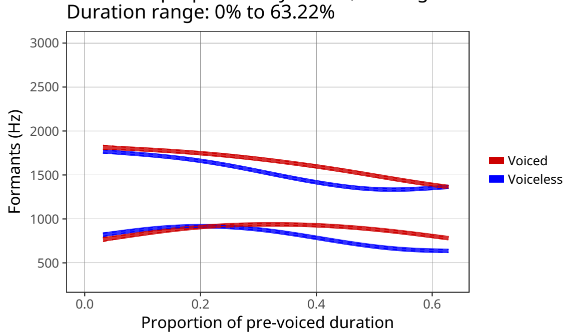


aw

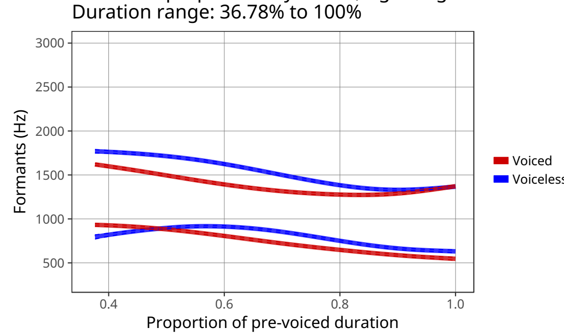
SSANOVA of /aw/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



SSANOVA of /aw/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 63.22%

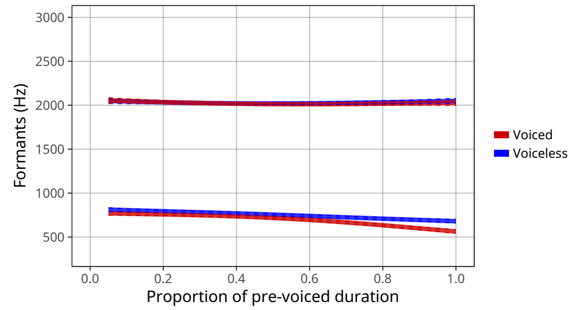


SSANOVA of /aw/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 36.78% to 100%

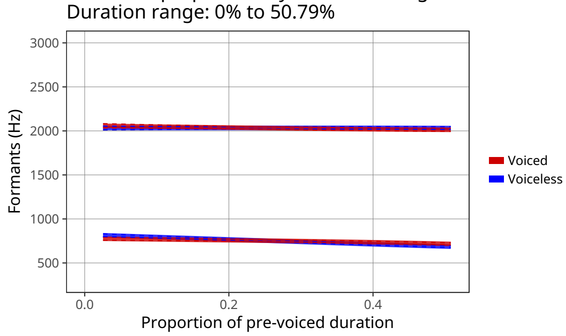


ɛ

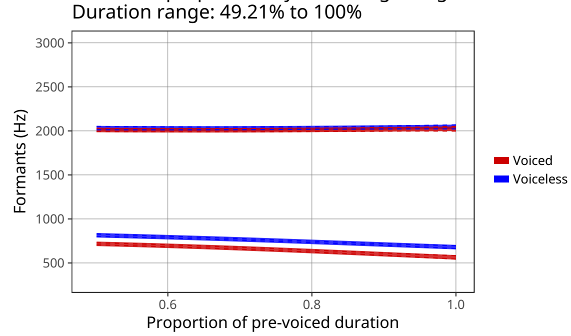
SSANOVA of /ɛ/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



SSANOVA of /ɛ/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 50.79%

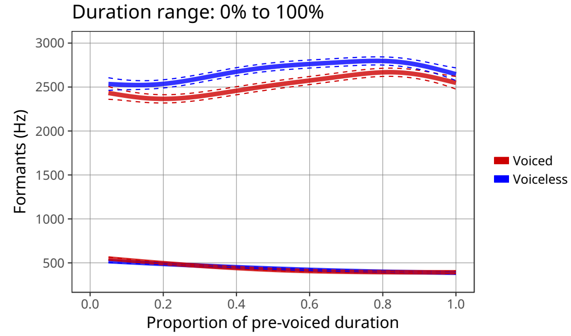


SSANOVA of /ɛ/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 49.21% to 100%

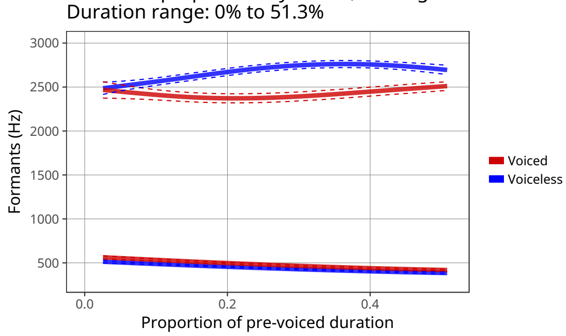


e

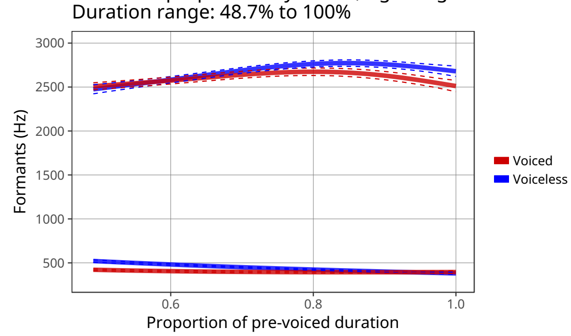
SSANOVA of /e/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



SSANOVA of /e/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 51.3%

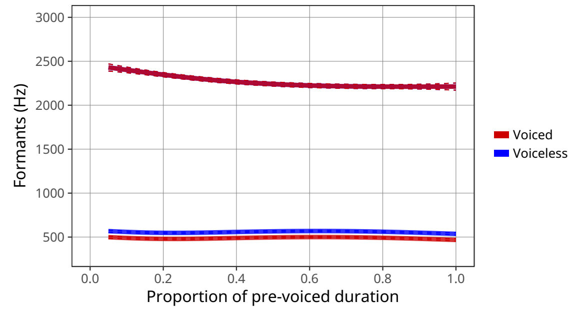


SSANOVA of /e/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 48.7% to 100%

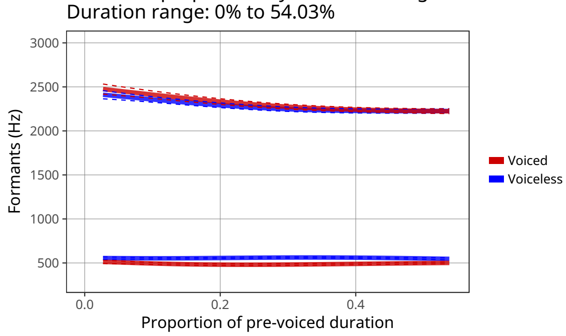


ɪ

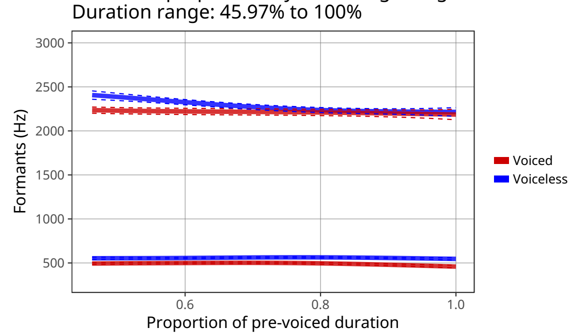
SSANOVA of /ɪ/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



SSANOVA of /ɪ/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 54.03%

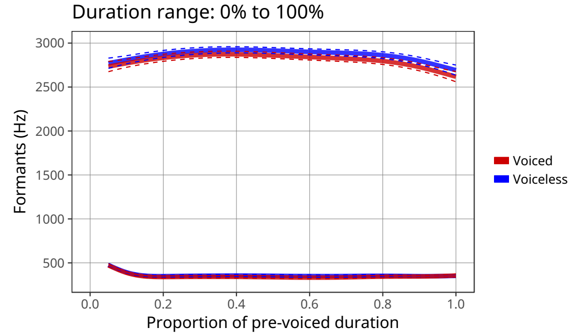


SSANOVA of /ɪ/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 45.97% to 100%

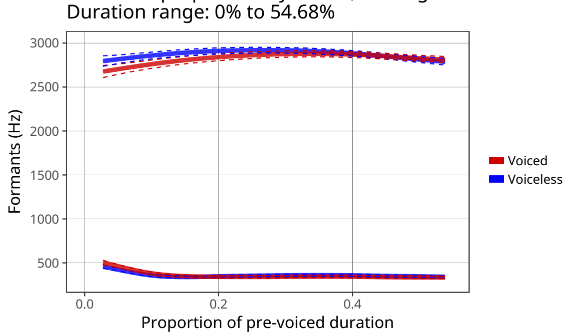


i

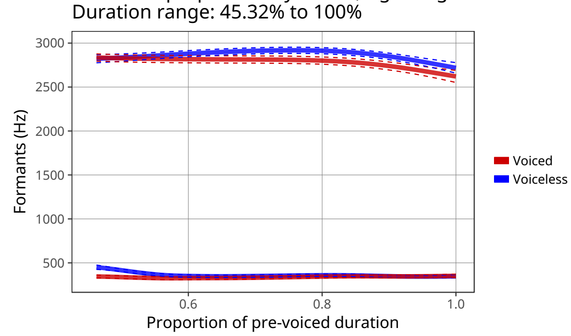
SSANOVA of /i/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



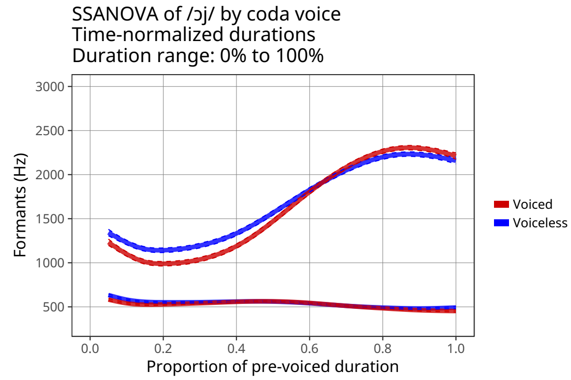
SSANOVA of /i/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 54.68%

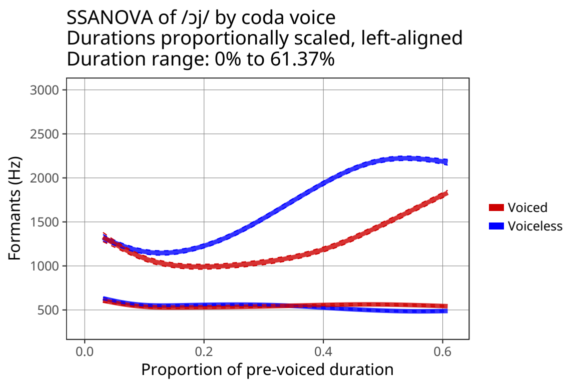


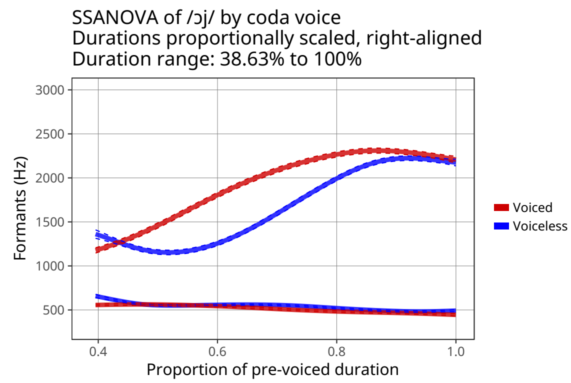
SSANOVA of /i/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 45.32% to 100%



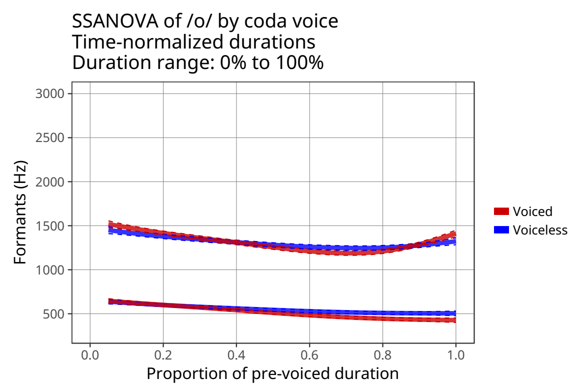
ɔj

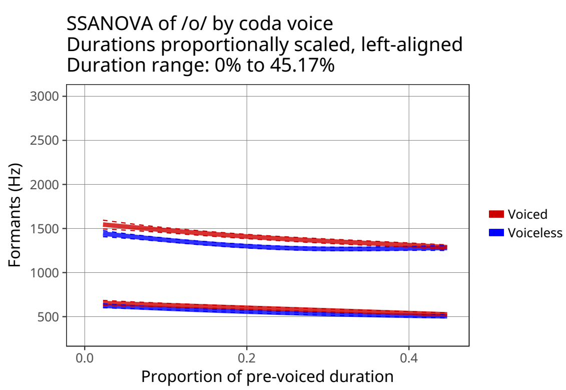

SSANOVA of /ɔj/ by coda voice
Time-normalized durations
Duration range: 0% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

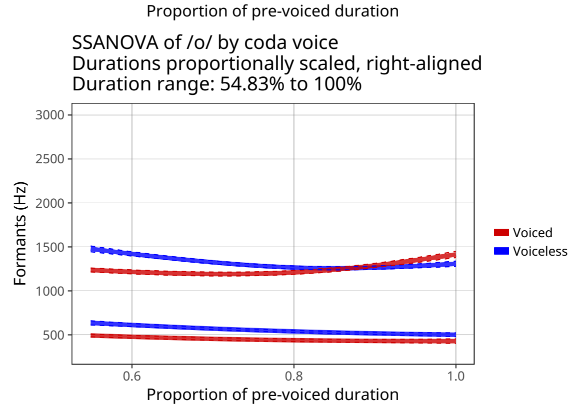

SSANOVA of /ɔj/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 61.37%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless


SSANOVA of /ɔj/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 38.63% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

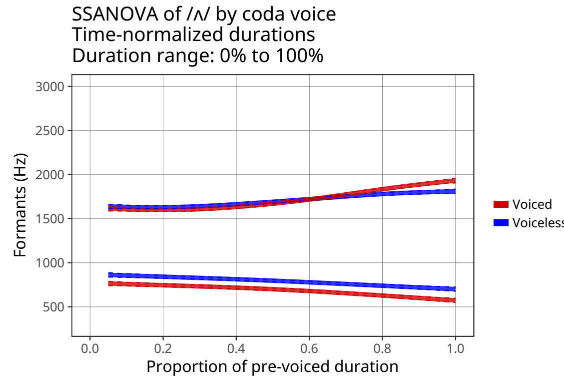
o

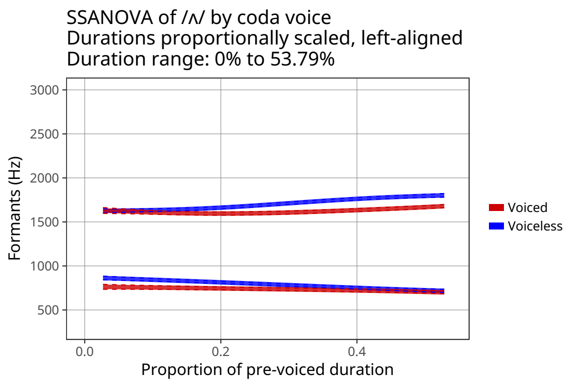

SSANOVA of /o/ by coda voice
Time-normalized durations
Duration range: 0% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

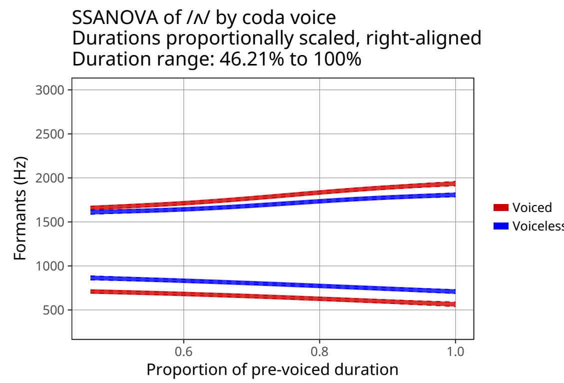

SSANOVA of /o/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 45.17%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless


SSANOVA of /o/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 54.83% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

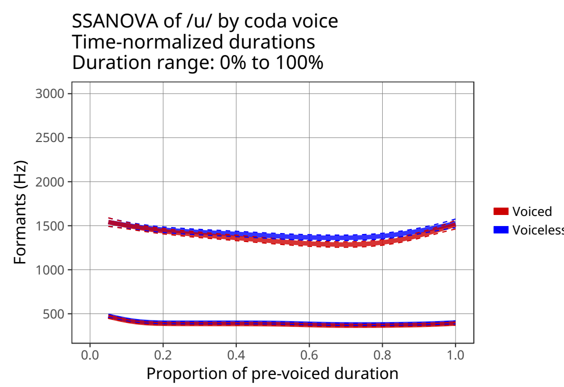
ʌ

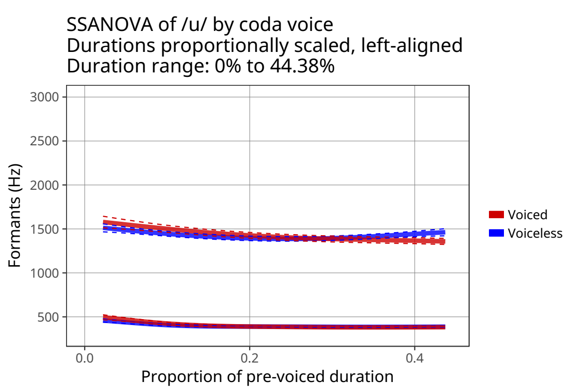

SSANOVA of /ʌ/ by coda voice
Time-normalized durations
Duration range: 0% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

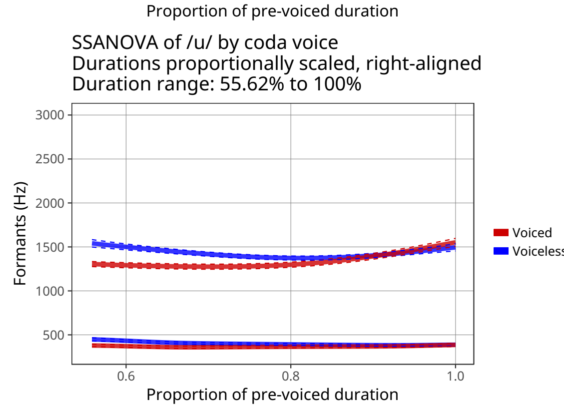

SSANOVA of /ʌ/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 53.79%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless


SSANOVA of /ʌ/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 46.21% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

u


SSANOVA of /u/ by coda voice
Time-normalized durations
Duration range: 0% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless


SSANOVA of /u/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 44.38%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

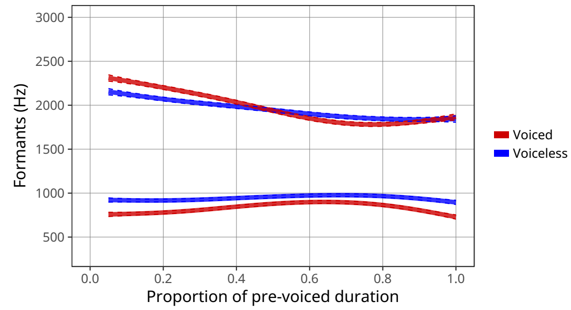

SSANOVA of /u/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 55.62% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

Appendix D SSANOVA comparisons, Madison

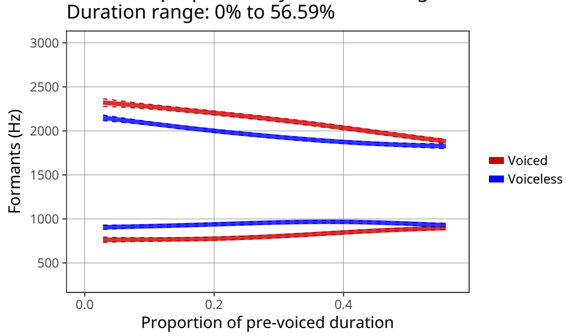
This appendix contains the full set of SSANOVA comparisons for each durational scaling /alignment model (see §4.4.2) across all vowels in the Madison dataset. The time-normalized comparison model is listed at the top for each vowel, followed by the proportionally-scaled duration comparison models below, with left- and right-alignments, respectively.

æ

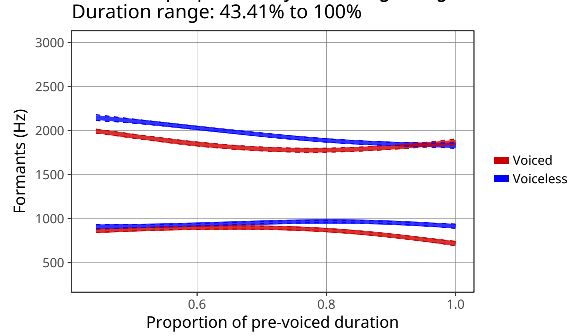
SSANOVA of /æ/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



SSANOVA of /æ/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 56.59%

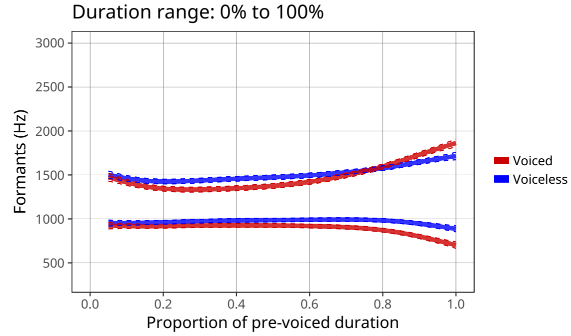


SSANOVA of /æ/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 43.41% to 100%

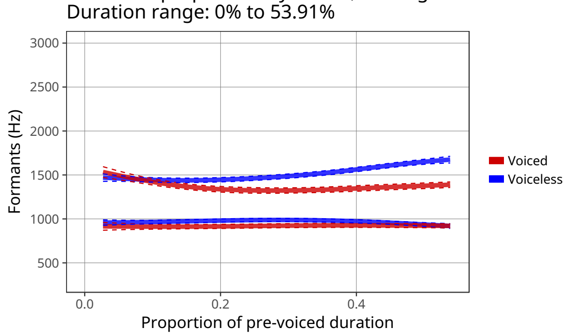


ɑ

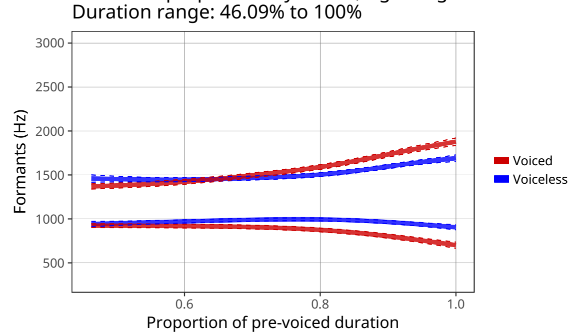
SSANOVA of /ɑ/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



SSANOVA of /ɑ/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 53.91%

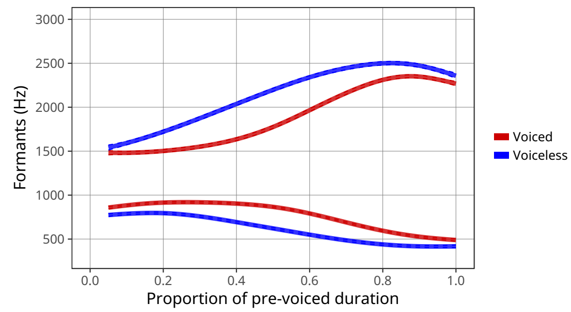


SSANOVA of /ɑ/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 46.09% to 100%

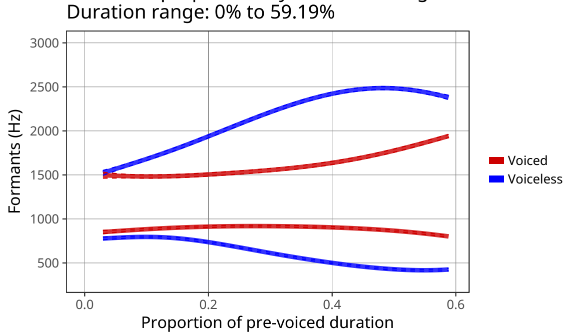


aj

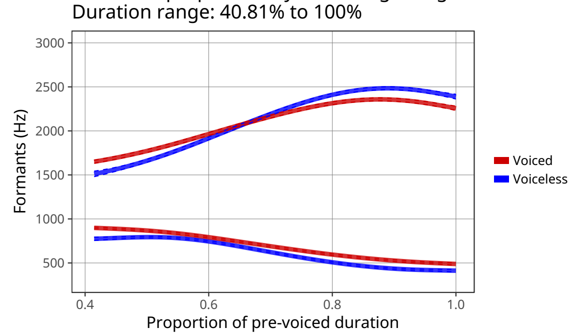
SSANOVA of /aj/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



SSANOVA of /aj/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 59.19%

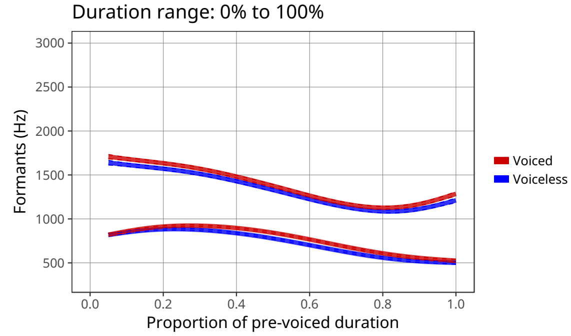


SSANOVA of /aj/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 40.81% to 100%

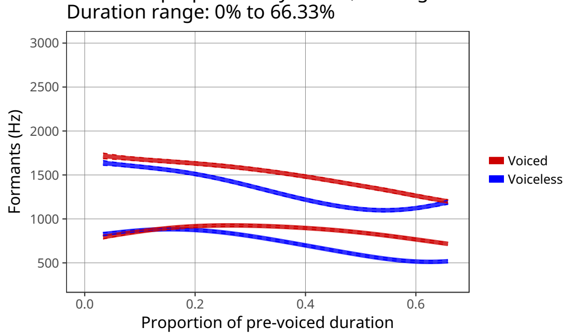


aw

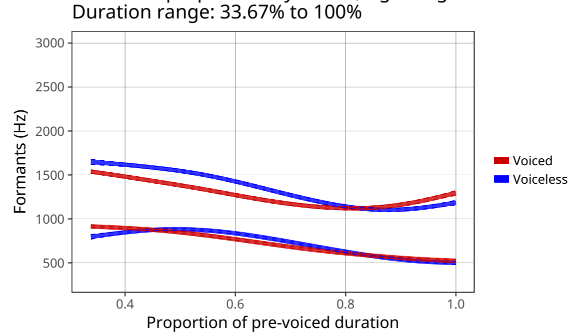
SSANOVA of /aw/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



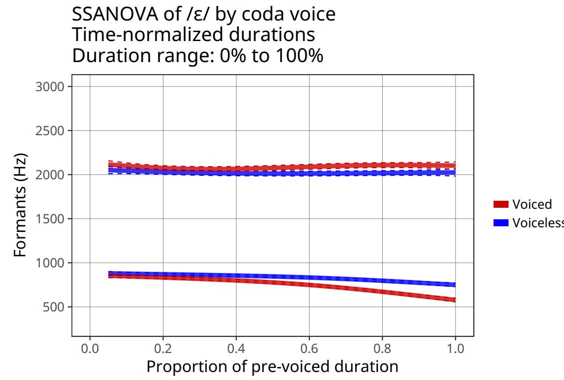
SSANOVA of /aw/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 66.33%

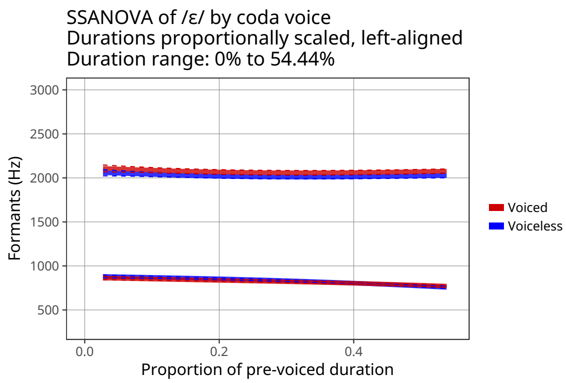


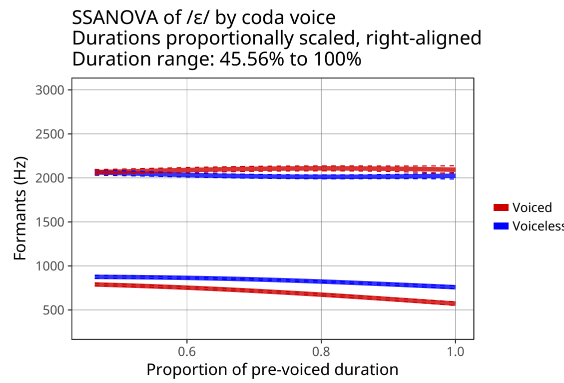
SSANOVA of /aw/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 33.67% to 100%



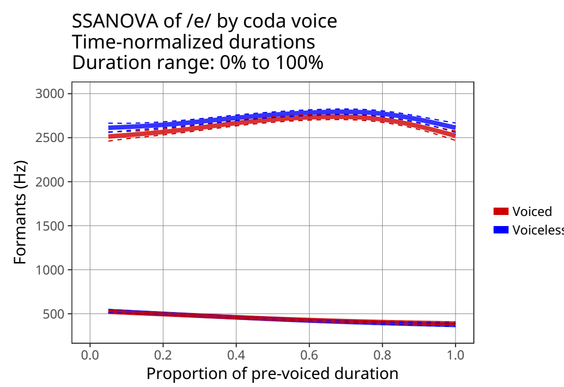
ɛ

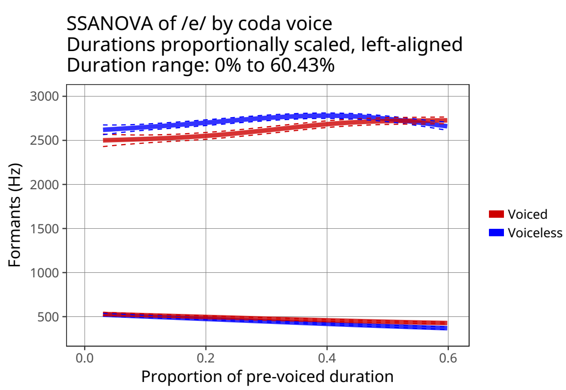

SSANOVA of /ɛ/ by coda voice
Time-normalized durations
Duration range: 0% to 100%
Formants (Hz)
3000
2500
2000
1500
1000
500
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of pre-voiced duration
Voiced
Voiceless

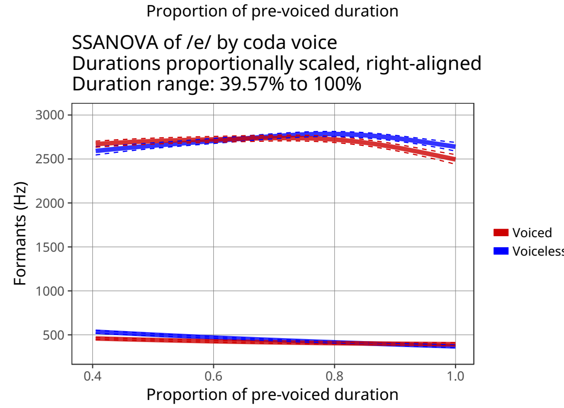

SSANOVA of /ɛ/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 54.44%
Formants (Hz)
3000
2500
2000
1500
1000
500
0.0
0.2
0.4
Proportion of pre-voiced duration
Voiced
Voiceless


SSANOVA of /ɛ/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 45.56% to 100%
Formants (Hz)
3000
2500
2000
1500
1000
500
0.6
0.8
1.0
Proportion of pre-voiced duration
Voiced
Voiceless

e

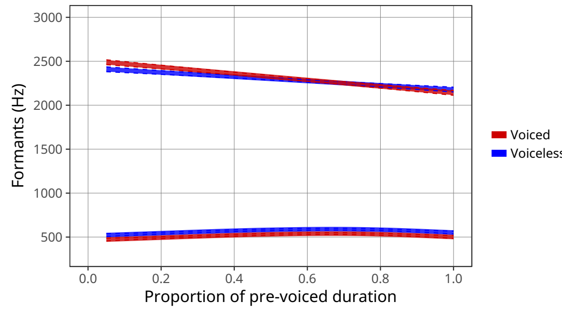

SSANOVA of /e/ by coda voice
Time-normalized durations
Duration range: 0% to 100%
Formants (Hz)
3000
2500
2000
1500
1000
500
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of pre-voiced duration
Voiced
Voiceless


SSANOVA of /e/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 60.43%
Formants (Hz)
3000
2500
2000
1500
1000
500
0.0
0.2
0.4
0.6
Proportion of pre-voiced duration
Voiced
Voiceless

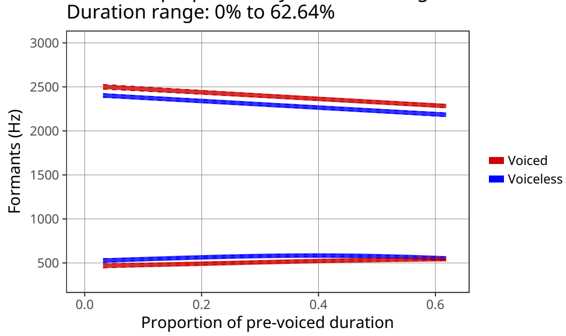

SSANOVA of /e/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 39.57% to 100%
Formants (Hz)
3000
2500
2000
1500
1000
500
0.4
0.6
0.8
1.0
Proportion of pre-voiced duration
Voiced
Voiceless

ɪ

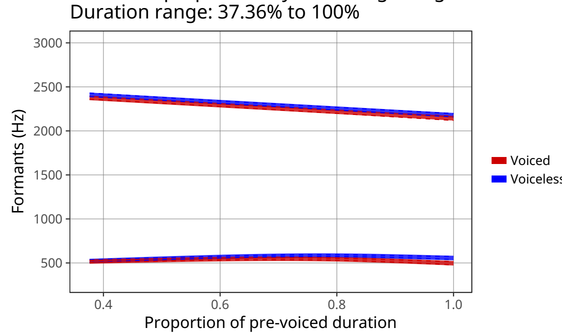
SSANOVA of /ɪ/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



SSANOVA of /ɪ/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 62.64%

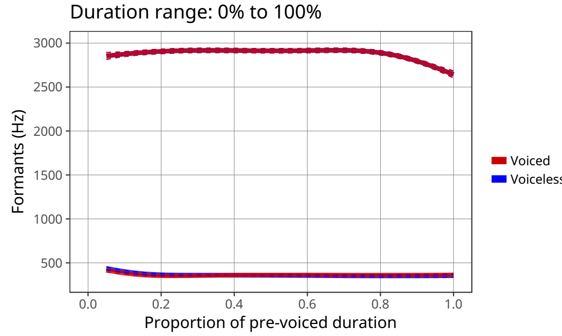


SSANOVA of /ɪ/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 37.36% to 100%

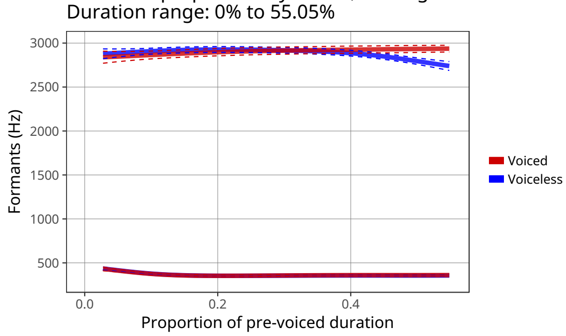


i

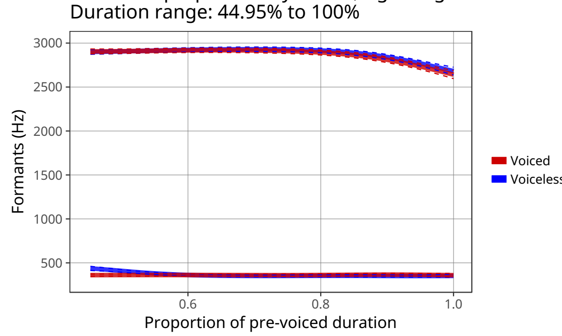
SSANOVA of /i/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



SSANOVA of /i/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 55.05%

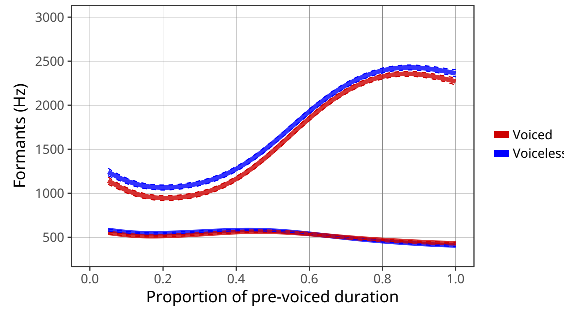


SSANOVA of /i/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 44.95% to 100%

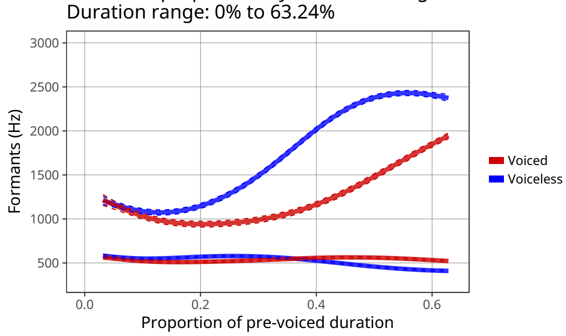


ɔj

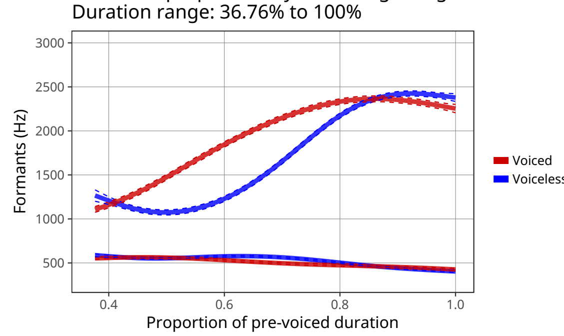
SSANOVA of /ɔj/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



SSANOVA of /ɔj/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 63.24%

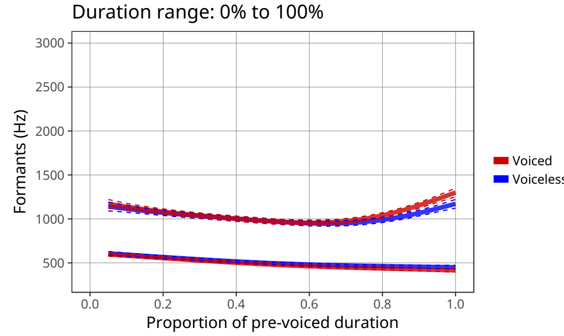


SSANOVA of /ɔj/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 36.76% to 100%

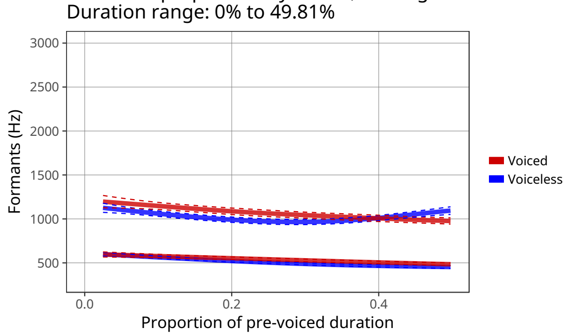


o

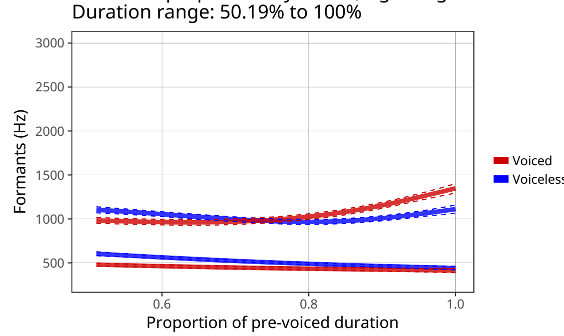
SSANOVA of /o/ by coda voice
Time-normalized durations
Duration range: 0% to 100%



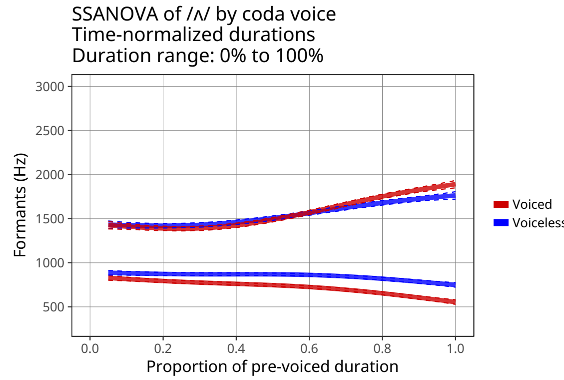
SSANOVA of /o/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 49.81%

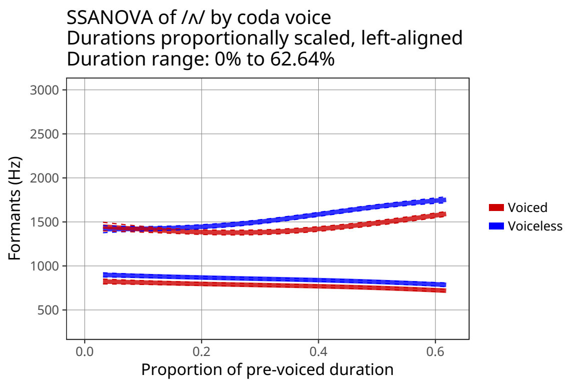


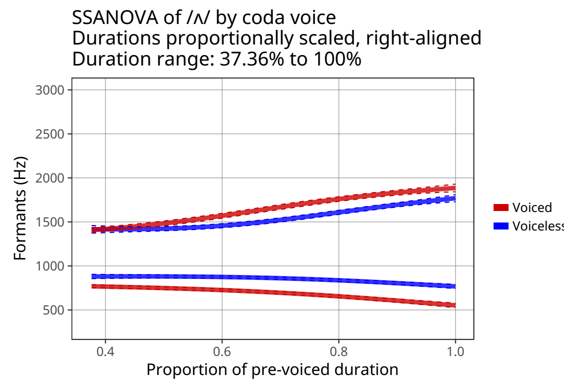
SSANOVA of /o/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 50.19% to 100%



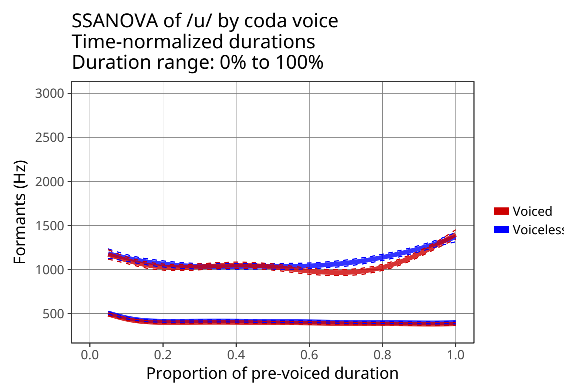
ʌ

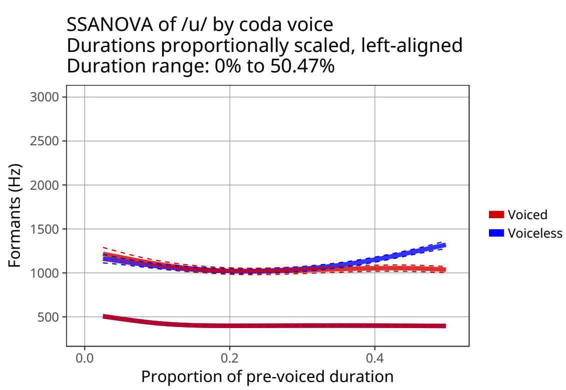

SSANOVA of /ʌ/ by coda voice
Time-normalized durations
Duration range: 0% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

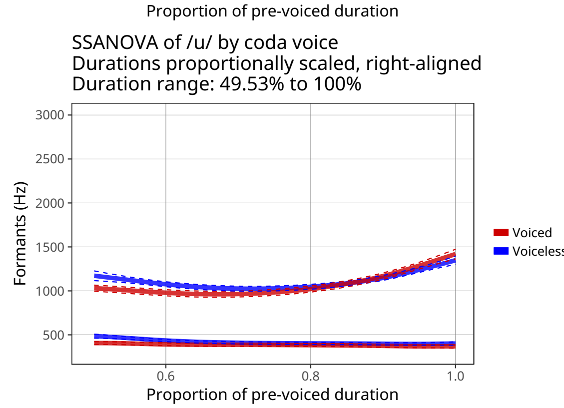

SSANOVA of /ʌ/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 62.64%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless


SSANOVA of /ʌ/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 37.36% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

u


SSANOVA of /u/ by coda voice
Time-normalized durations
Duration range: 0% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

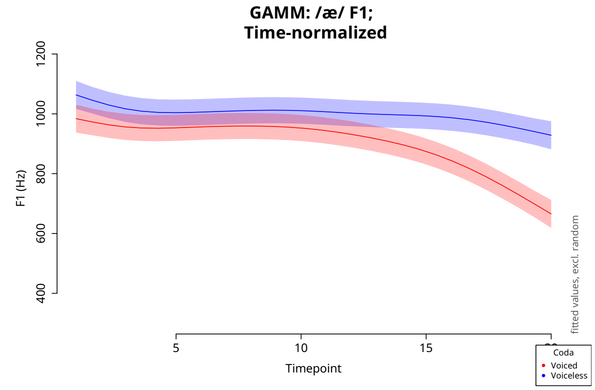

SSANOVA of /u/ by coda voice
Durations proportionally scaled, left-aligned
Duration range: 0% to 50.47%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

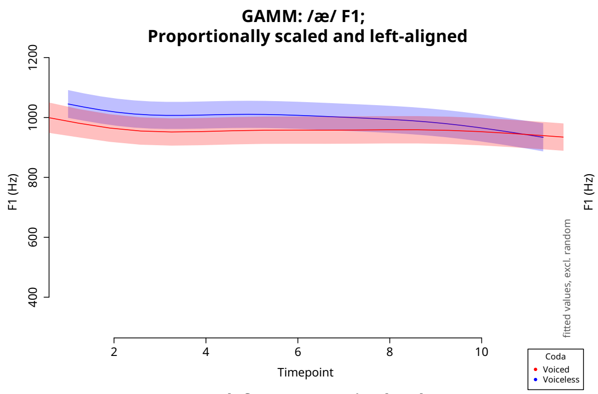

SSANOVA of /u/ by coda voice
Durations proportionally scaled, right-aligned
Duration range: 49.53% to 100%
Formants (Hz)
Proportion of pre-voiced duration
Voiced
Voiceless

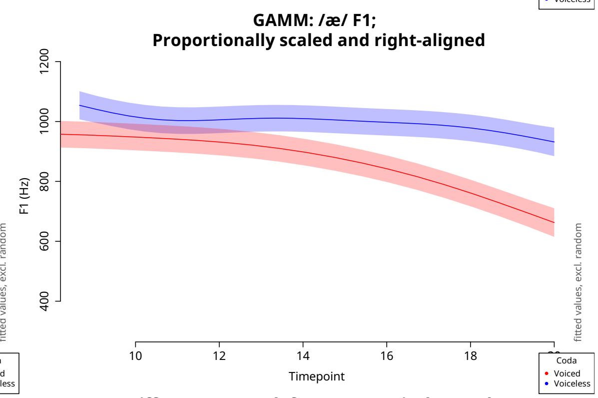
Appendix E GAMMs comparisons, Winnipeg

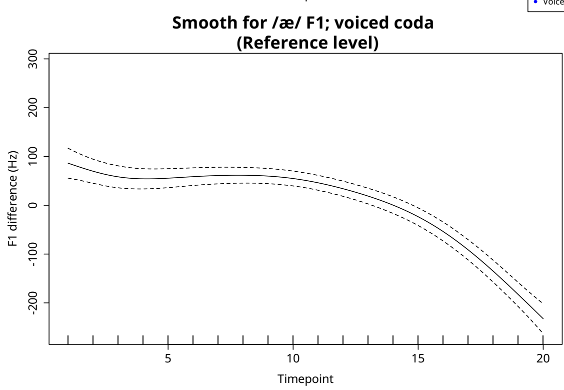
This appendix contains the full set of GAMMs comparisons for each durational scaling /alignment model (see §4.4.3) across all vowels in the Winnipeg dataset. The top set of three plots (in colour) display smooths with confidence intervals for each coda voicing condition for each model; the lower four plots are difference smooths, with the reference level (voiced coda) at top right, followed by the time-normalized model at top left, and the proportionally-scaled duration models, left- and right-alignment, in the lower left and right plots, respectively.

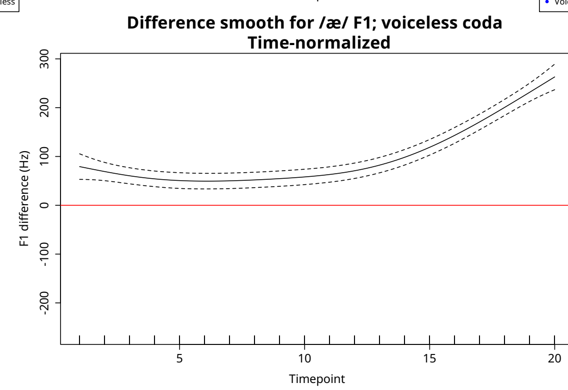
æ F1

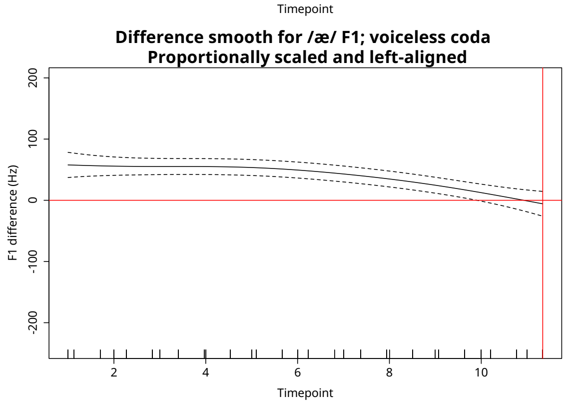

GAMM: /æ/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

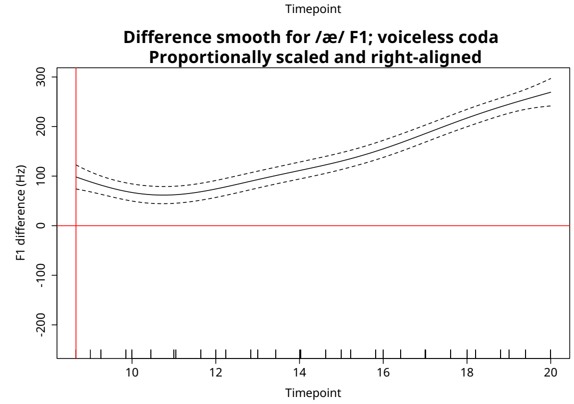

GAMM: /æ/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /æ/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


Smooth for /æ/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint

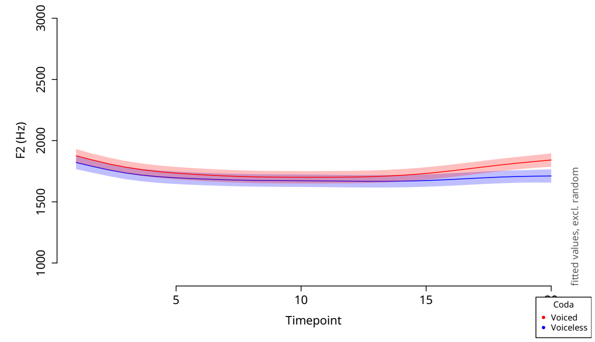

Difference smooth for /æ/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint


Difference smooth for /æ/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

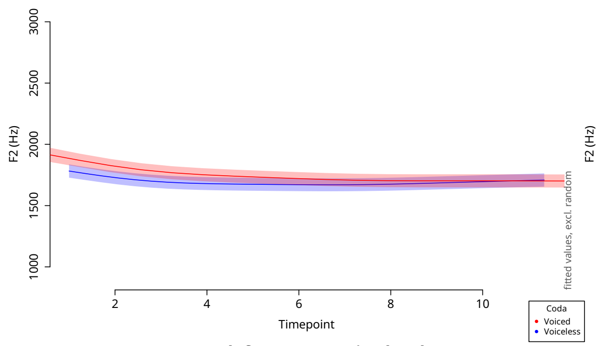

Difference smooth for /æ/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

æ F2

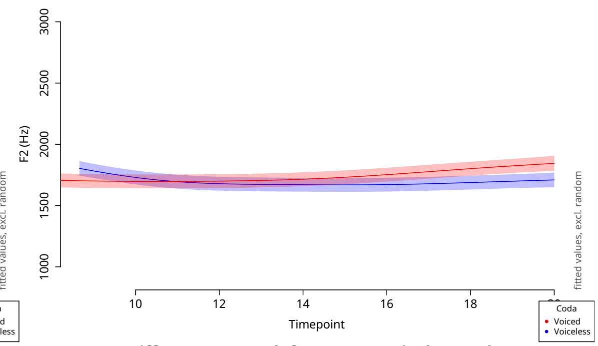
GAMM: /æ/ F2; Time-normalized



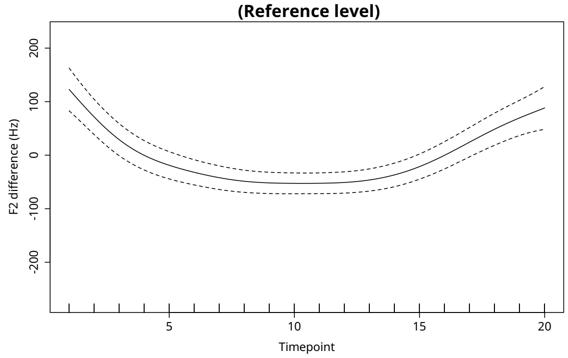
GAMM: /æ/ F2; Proportionally scaled and left-aligned



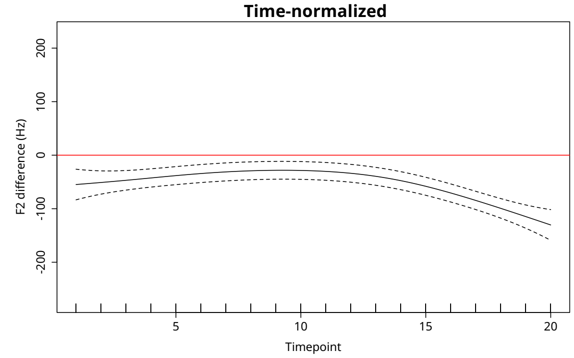
GAMM: /æ/ F2; Proportionally scaled and right-aligned



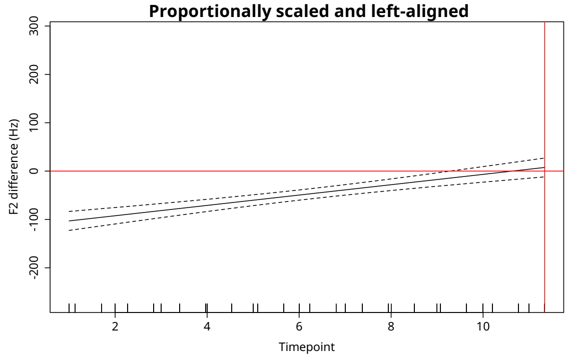
Smooth for /æ/ F2; voiced coda (Reference level)



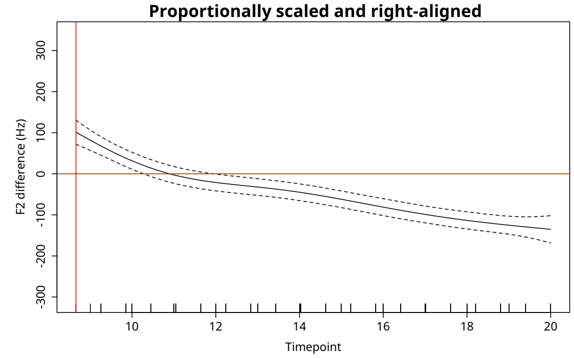
Difference smooth for /æ/ F2; voiceless coda Time-normalized



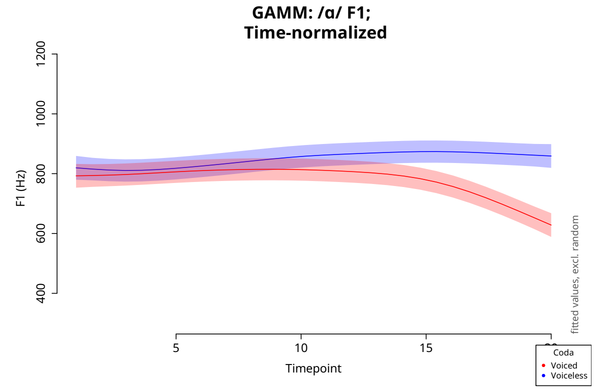
Difference smooth for /æ/ F2; voiceless coda Proportionally scaled and left-aligned



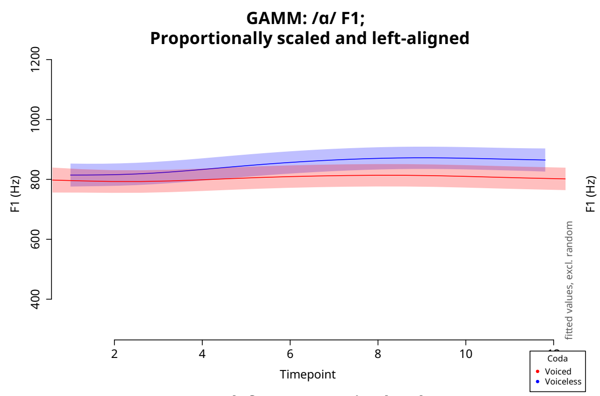
Difference smooth for /æ/ F2; voiceless coda Proportionally scaled and right-aligned



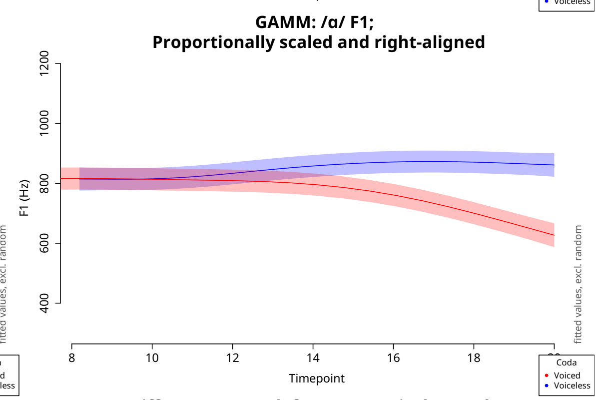
ɑ F1



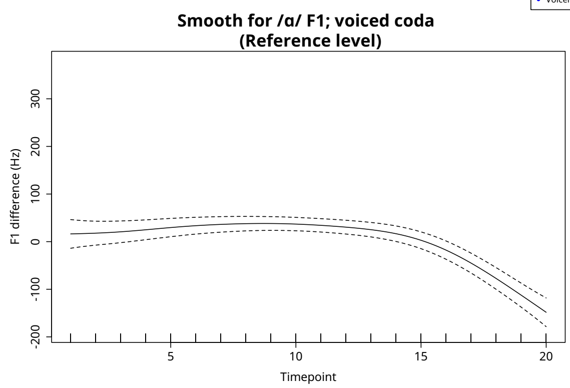
GAMM: /ɑ/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless



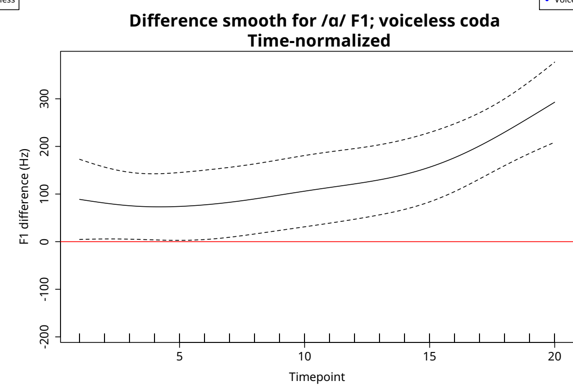
GAMM: /ɑ/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless



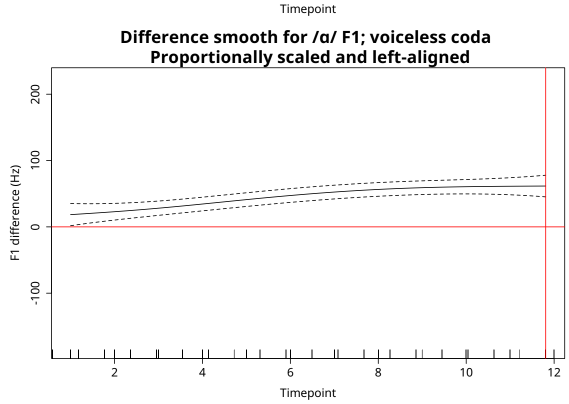
GAMM: /ɑ/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless



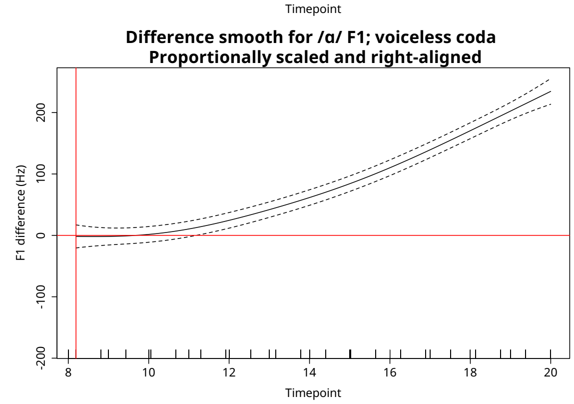
Smooth for /ɑ/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint



Difference smooth for /ɑ/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint

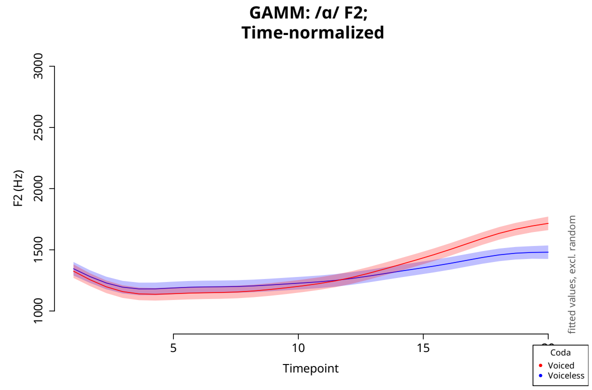


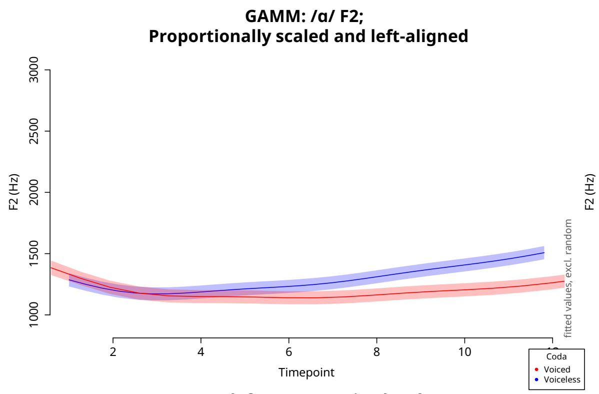
Difference smooth for /ɑ/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

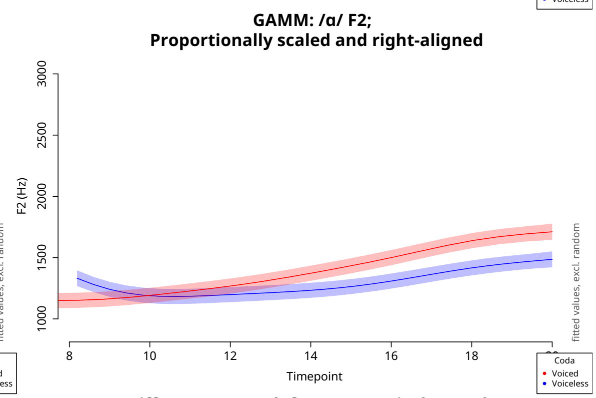


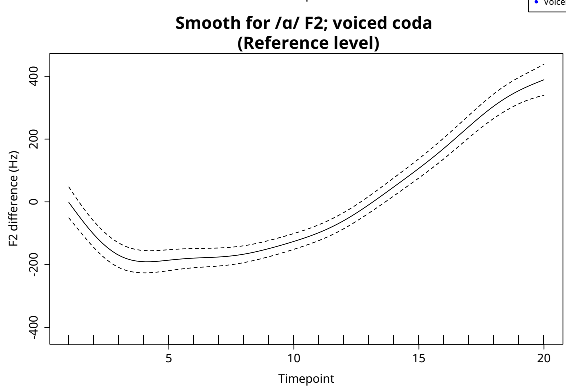
Difference smooth for /ɑ/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

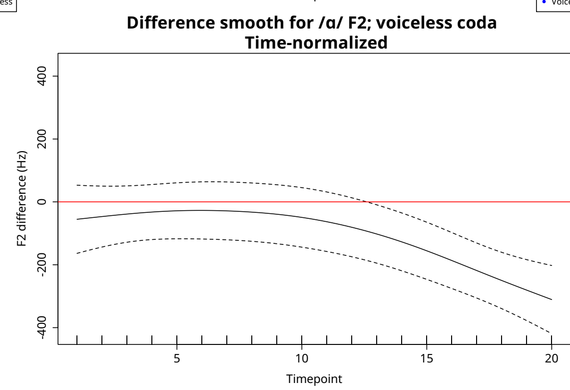
ɑ F2

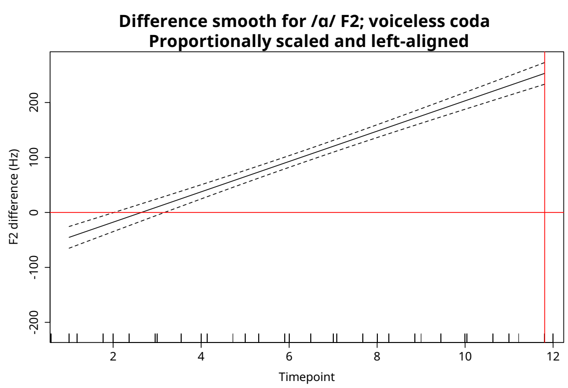

GAMM: /ɑ/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

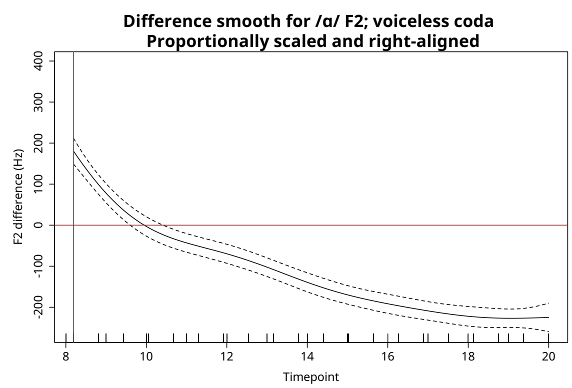

GAMM: /ɑ/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ɑ/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


Smooth for /ɑ/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint

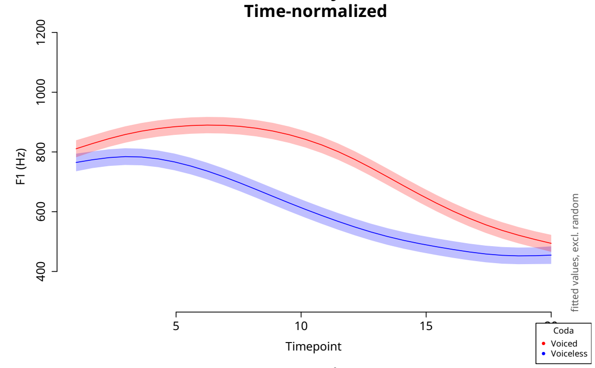

Difference smooth for /ɑ/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint


Difference smooth for /ɑ/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

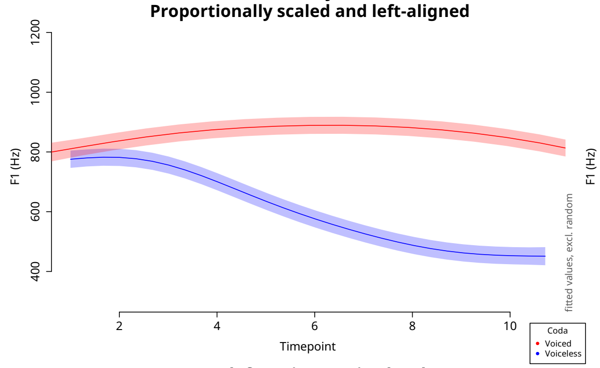

Difference smooth for /ɑ/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

aj F1

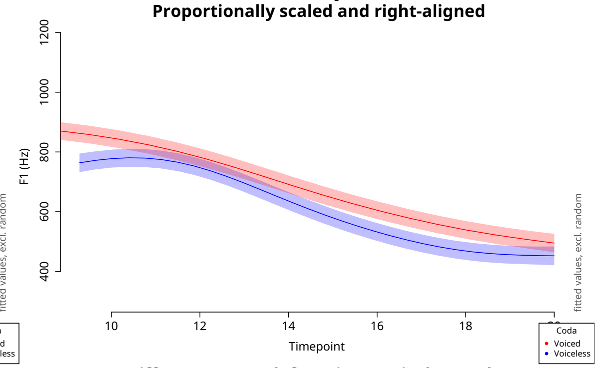
**GAMM: /aj/ F1;
Time-normalized**



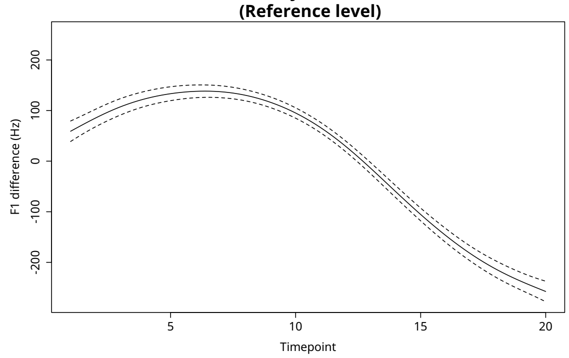
**GAMM: /aj/ F1;
Proportionally scaled and left-aligned**



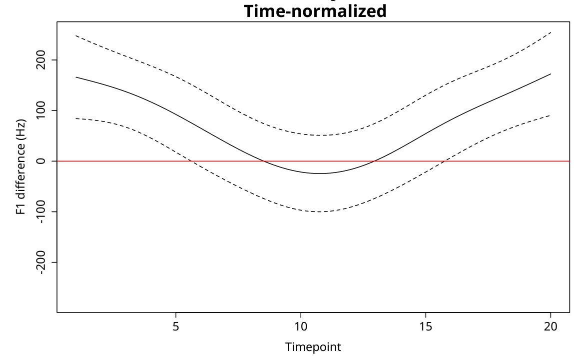
**GAMM: /aj/ F1;
Proportionally scaled and right-aligned**



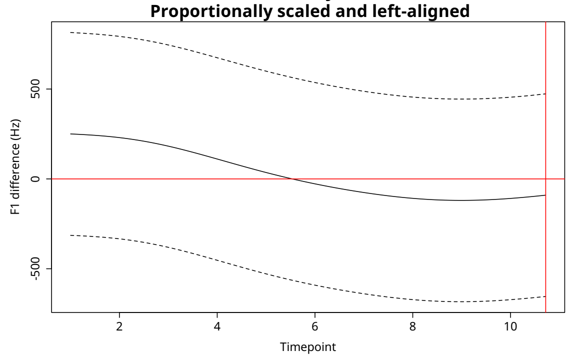
**Smooth for /aj/ F1; voiced coda
(Reference level)**



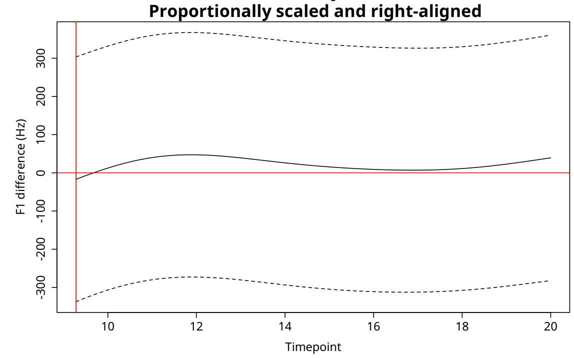
**Difference smooth for /aj/ F1; voiceless coda
Time-normalized**



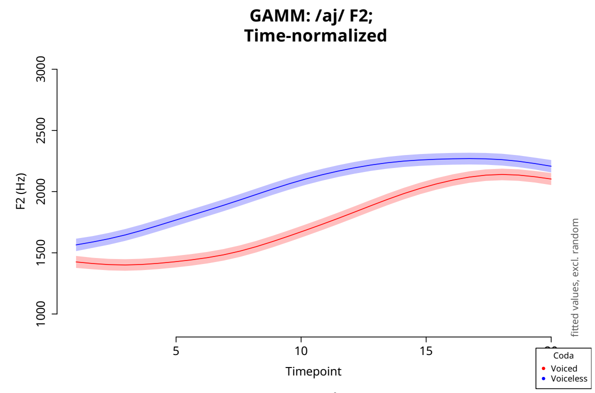
**Difference smooth for /aj/ F1; voiceless coda
Proportionally scaled and left-aligned**

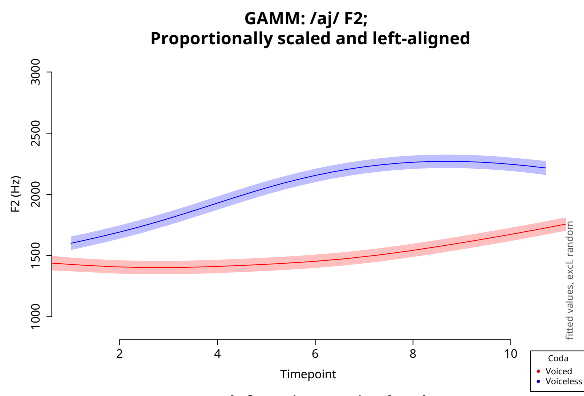


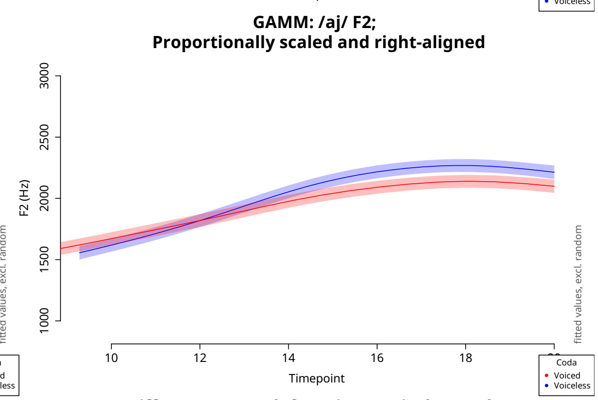
**Difference smooth for /aj/ F1; voiceless coda
Proportionally scaled and right-aligned**

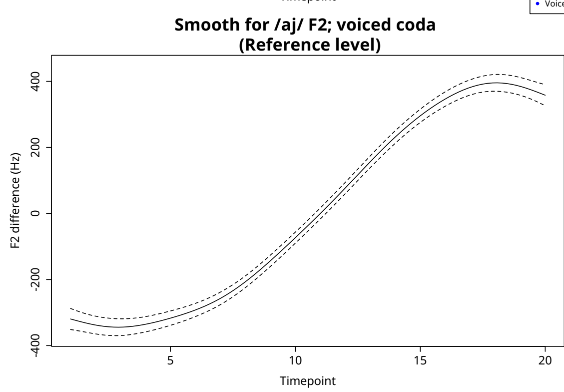


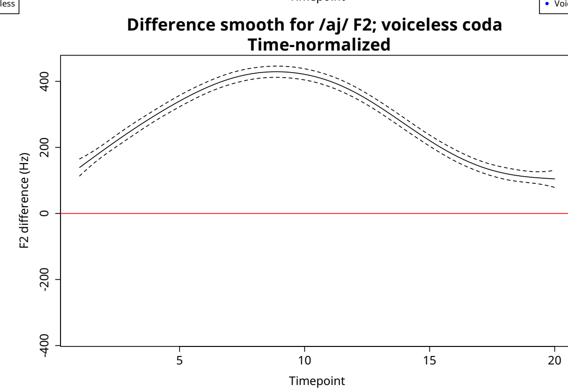
aj F2

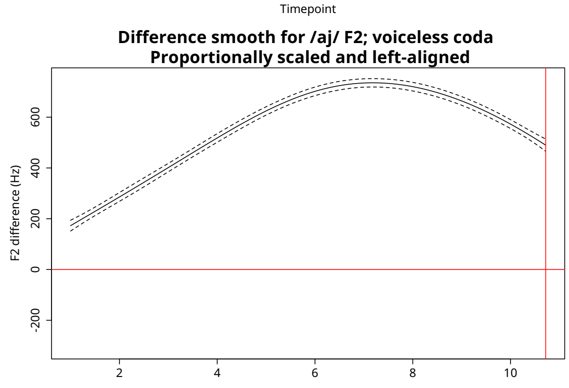

GAMM: /aj/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

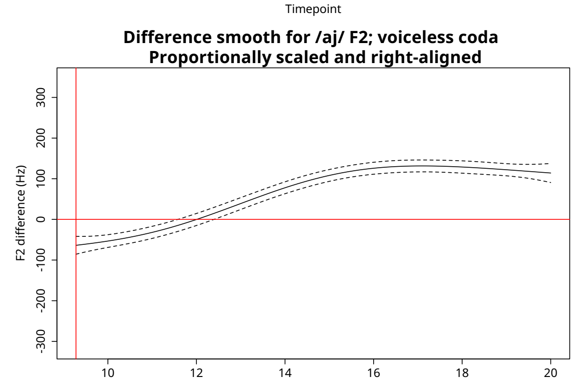

GAMM: /aj/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /aj/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


Smooth for /aj/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint

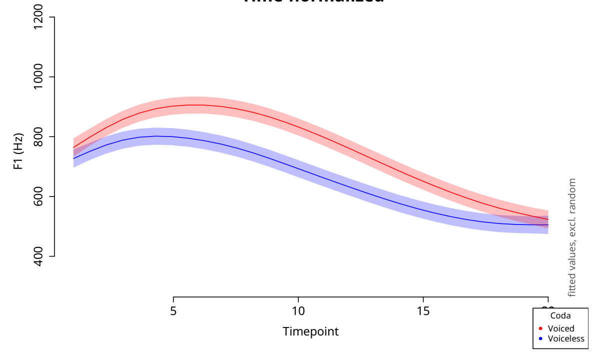

Difference smooth for /aj/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint


Difference smooth for /aj/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

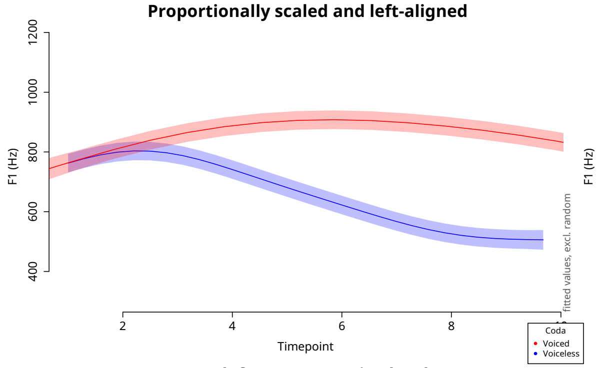

Difference smooth for /aj/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

aw F1

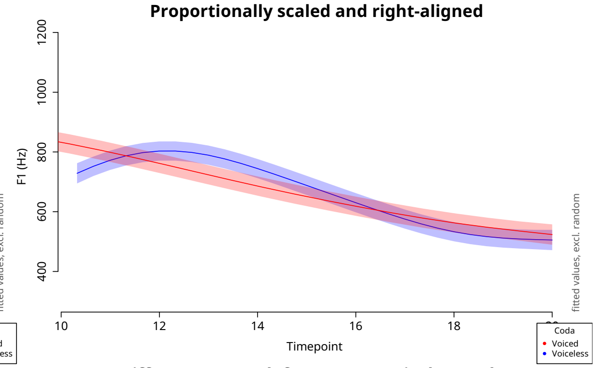
GAMM: /aw/ F1; Time-normalized



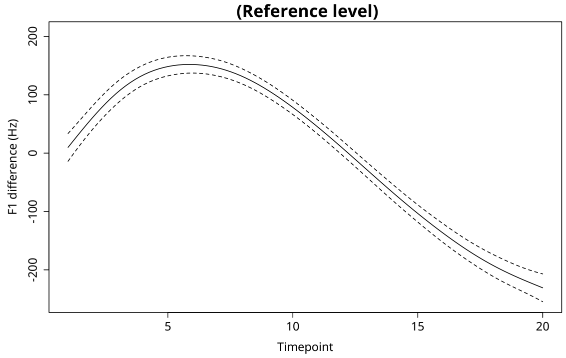
GAMM: /aw/ F1; Proportionally scaled and left-aligned



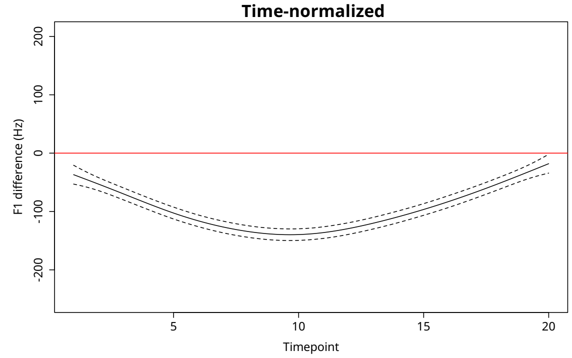
GAMM: /aw/ F1; Proportionally scaled and right-aligned



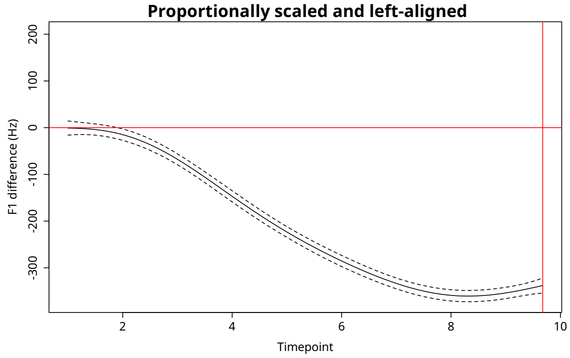
Smooth for /aw/ F1; voiced coda (Reference level)



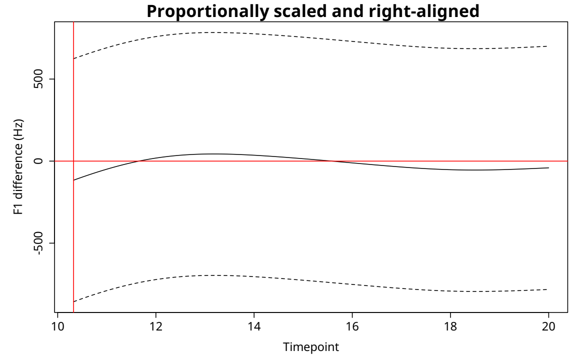
Difference smooth for /aw/ F1; voiceless coda Time-normalized



Difference smooth for /aw/ F1; voiceless coda Proportionally scaled and left-aligned

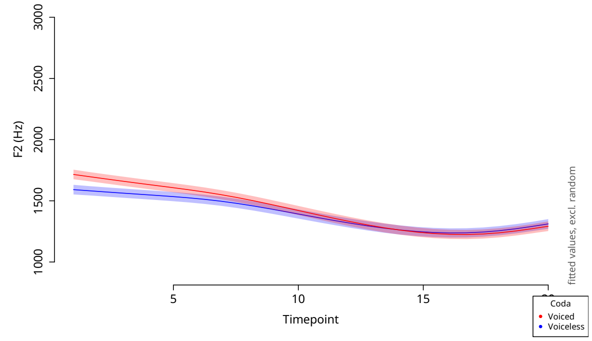


Difference smooth for /aw/ F1; voiceless coda Proportionally scaled and right-aligned

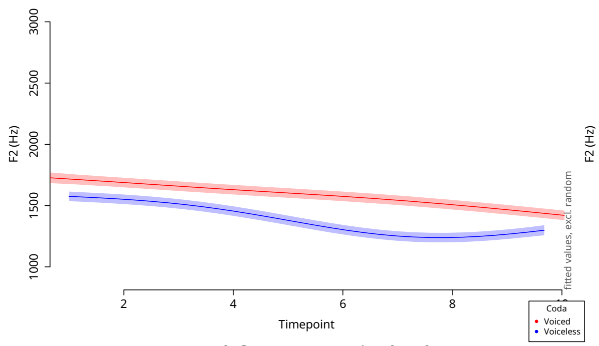


aw F2

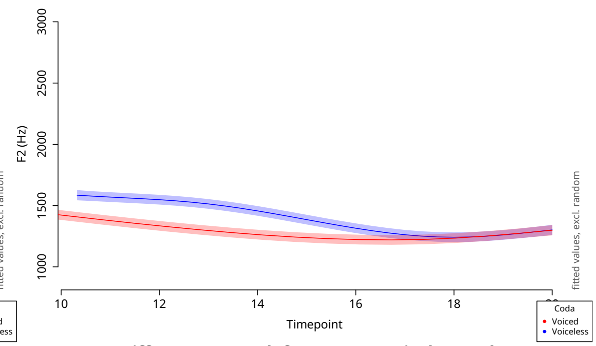
**GAMM: /aw/ F2;
Time-normalized**



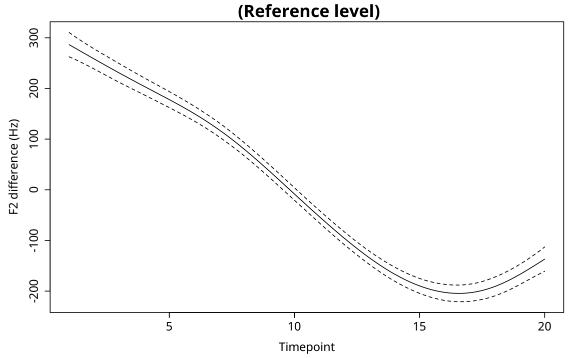
**GAMM: /aw/ F2;
Proportionally scaled and left-aligned**



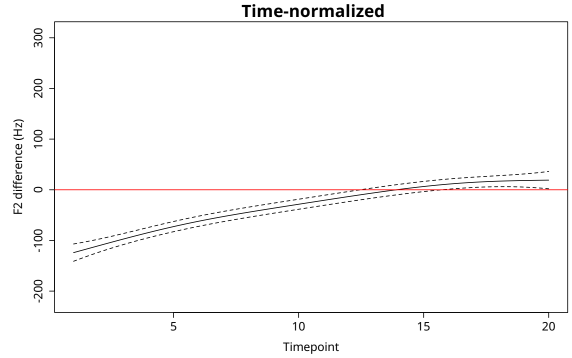
**GAMM: /aw/ F2;
Proportionally scaled and right-aligned**



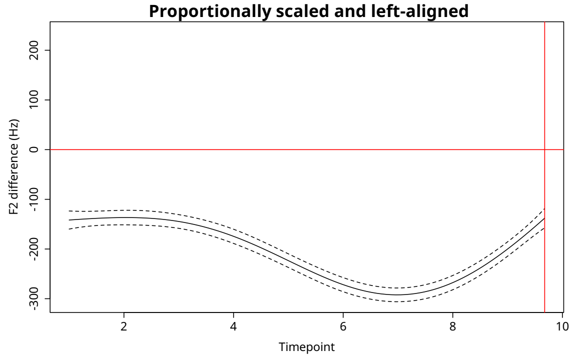
**Smooth for /aw/ F2; voiced coda
(Reference level)**



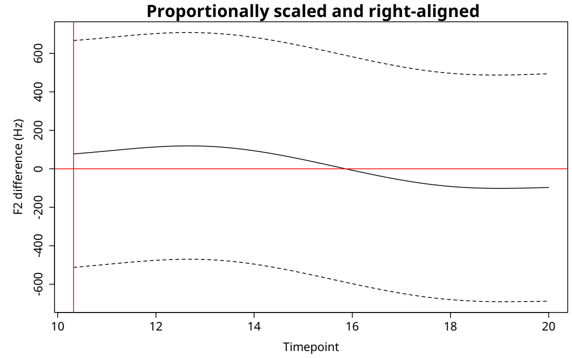
**Difference smooth for /aw/ F2; voiceless coda
Time-normalized**



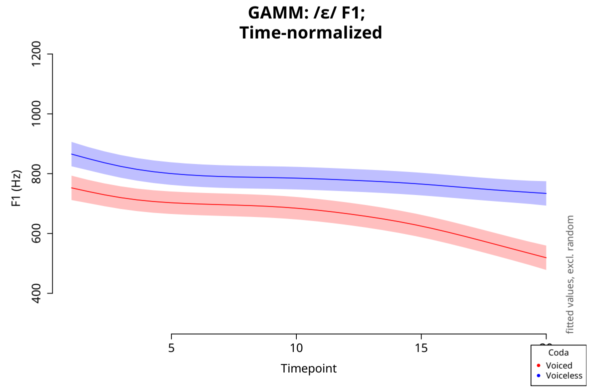
**Difference smooth for /aw/ F2; voiceless coda
Proportionally scaled and left-aligned**

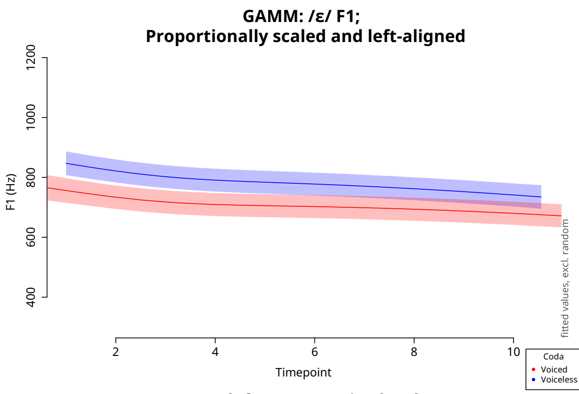


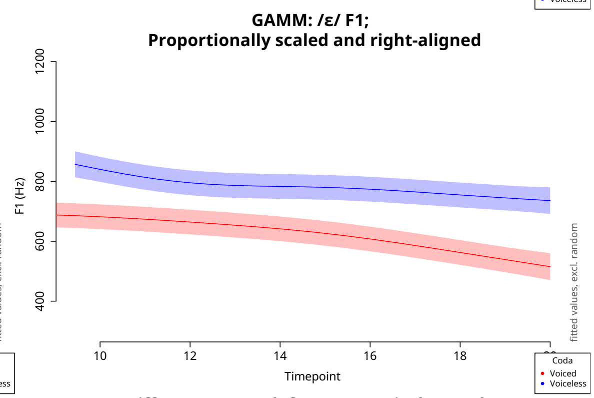
**Difference smooth for /aw/ F2; voiceless coda
Proportionally scaled and right-aligned**

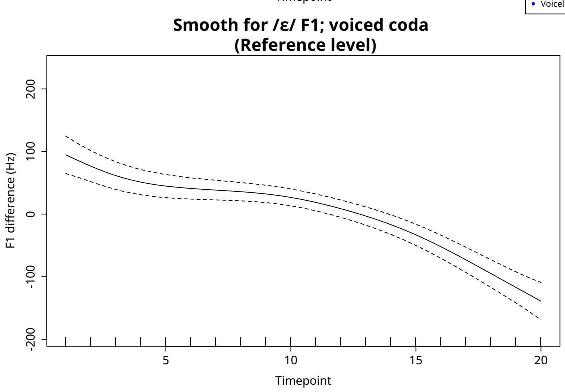


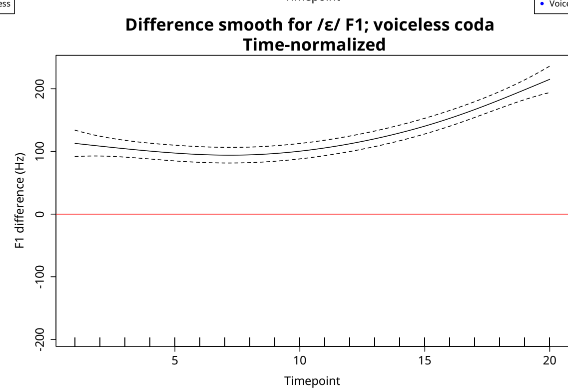
ɛ F1

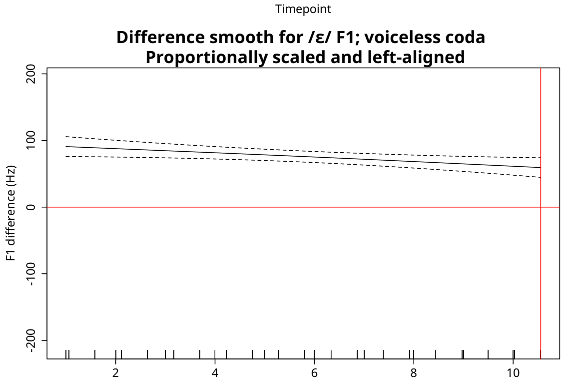

GAMM: /ɛ/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

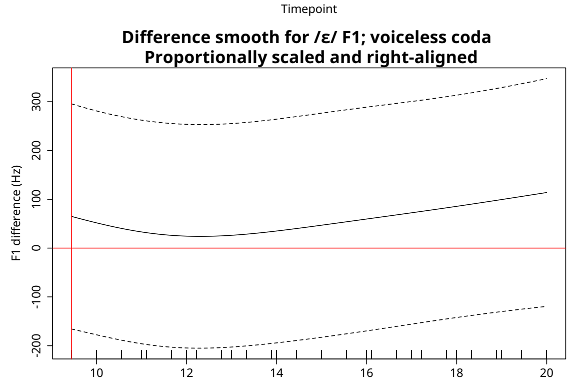

GAMM: /ɛ/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ɛ/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


Smooth for /ɛ/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint

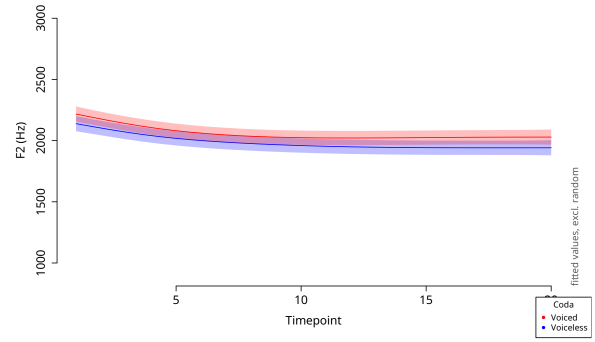

Difference smooth for /ɛ/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint


Difference smooth for /ɛ/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

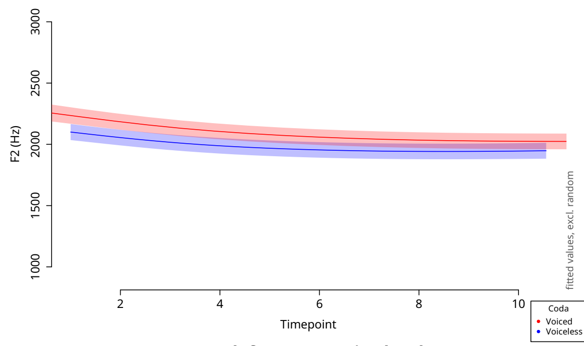

Difference smooth for /ɛ/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

ɛ F2

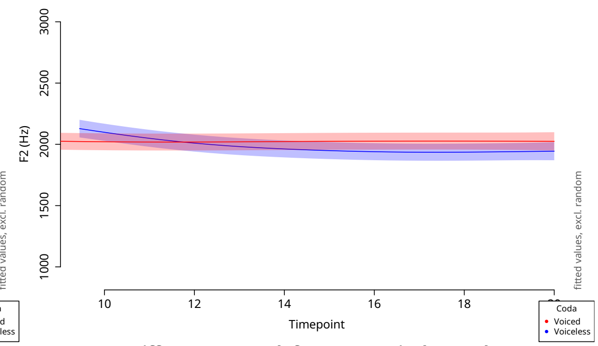
GAMM: /ɛ/ F2;
Time-normalized



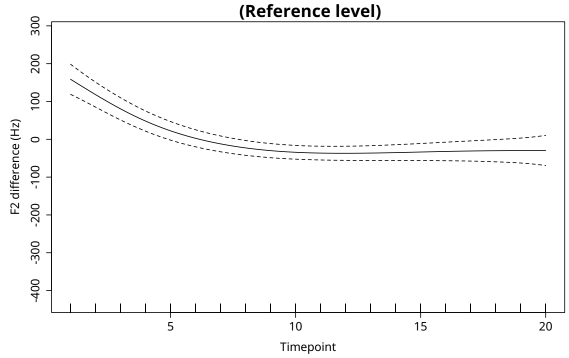
GAMM: /ɛ/ F2;
Proportionally scaled and left-aligned



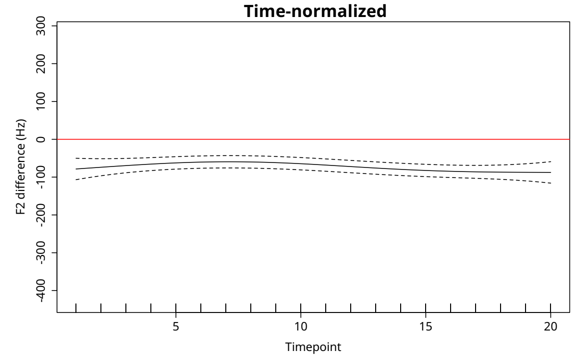
GAMM: /ɛ/ F2;
Proportionally scaled and right-aligned



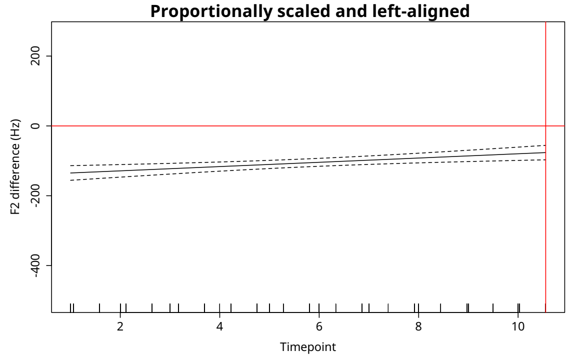
Smooth for /ɛ/ F2; voiced coda
(Reference level)



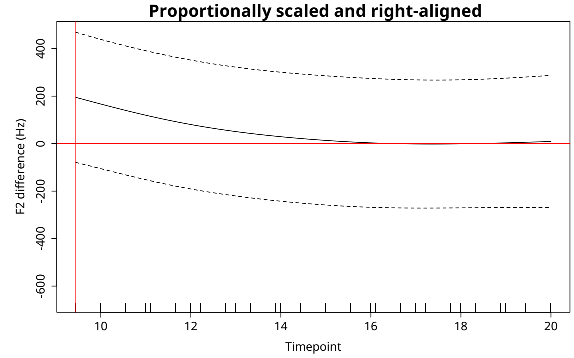
Difference smooth for /ɛ/ F2; voiceless coda
Time-normalized



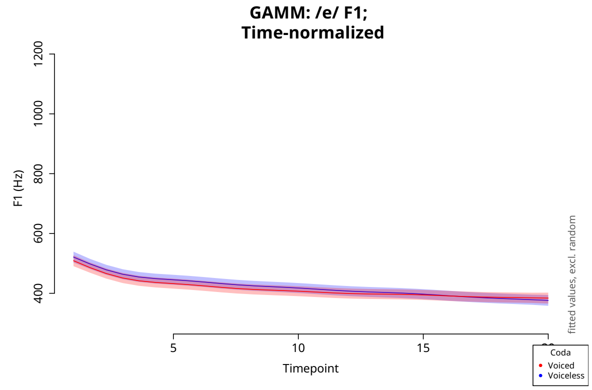
Difference smooth for /ɛ/ F2; voiceless coda
Proportionally scaled and left-aligned

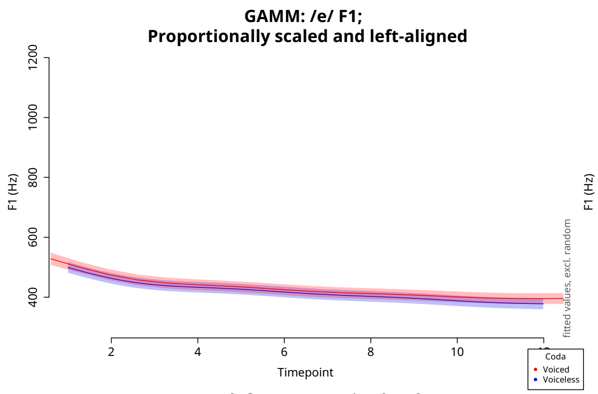


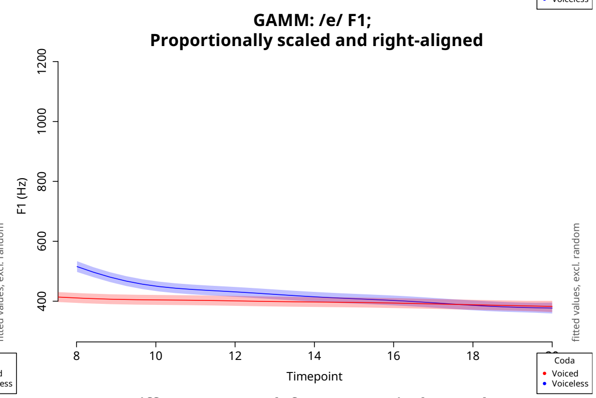
Difference smooth for /ɛ/ F2; voiceless coda
Proportionally scaled and right-aligned

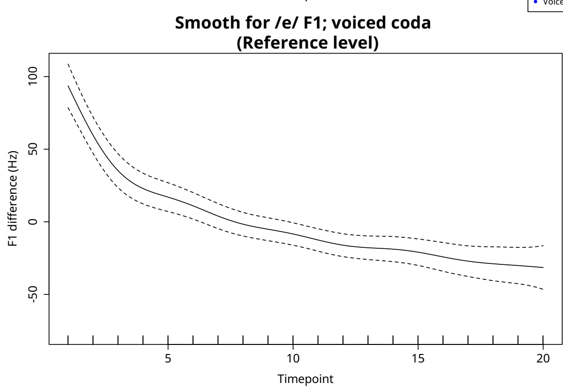


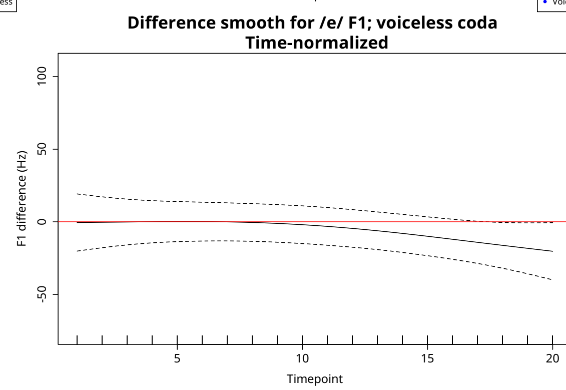
e F1

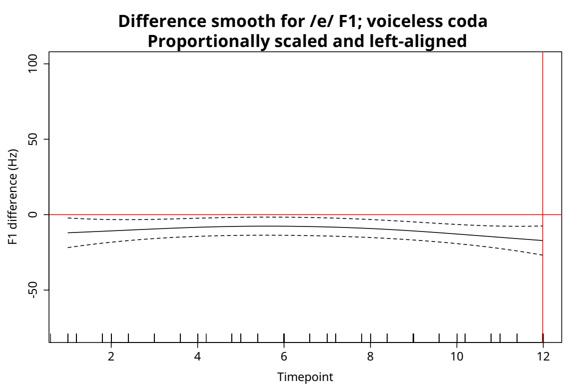

GAMM: /e/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

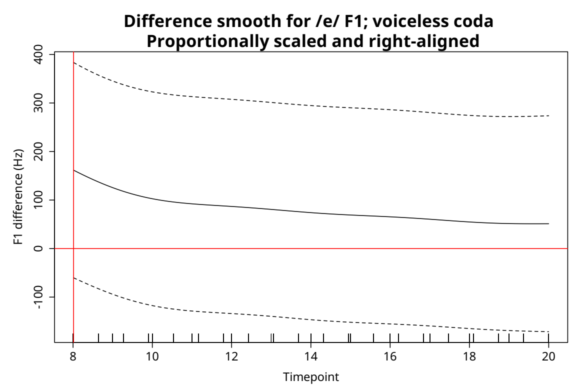

GAMM: /e/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /e/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

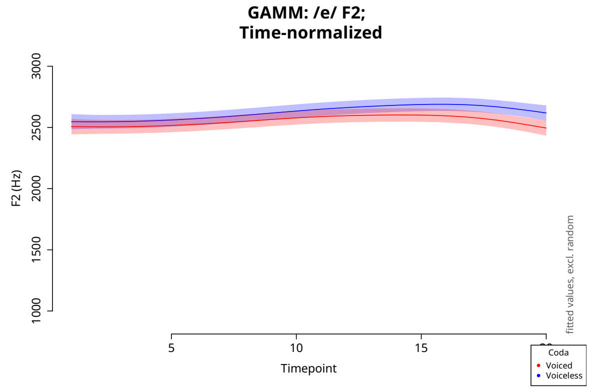

Smooth for /e/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint

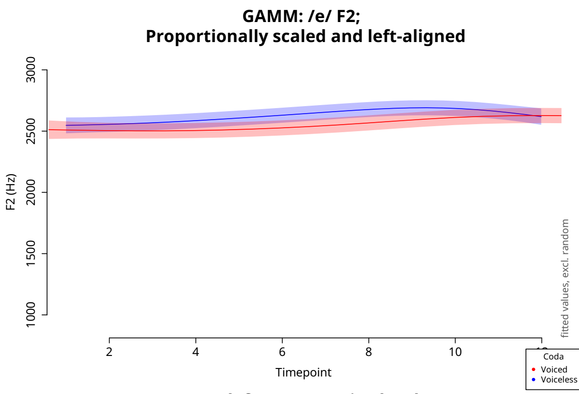

Difference smooth for /e/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint

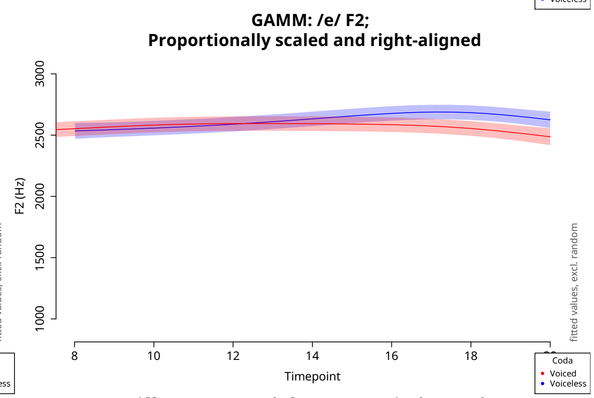

Difference smooth for /e/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

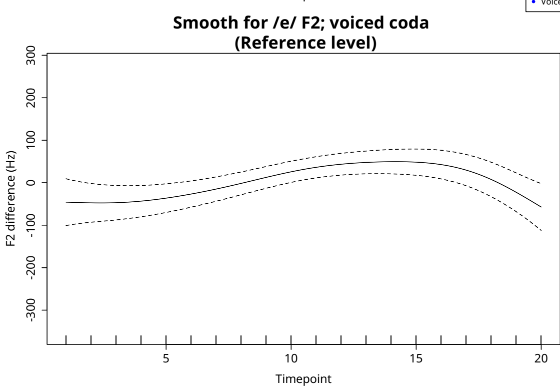

Difference smooth for /e/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

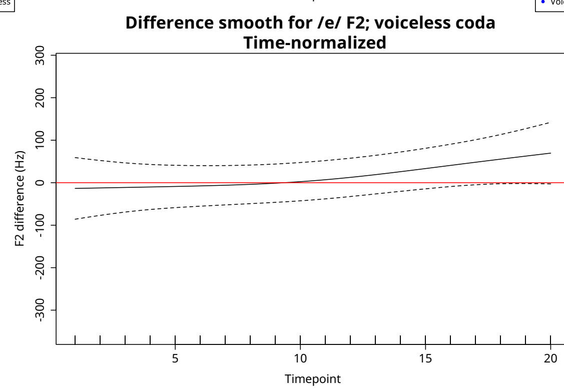
e F2

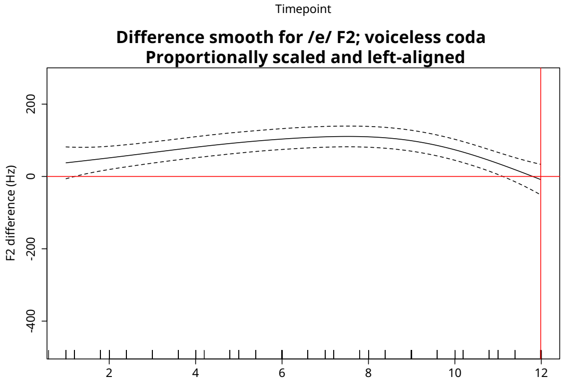

GAMM: /e/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

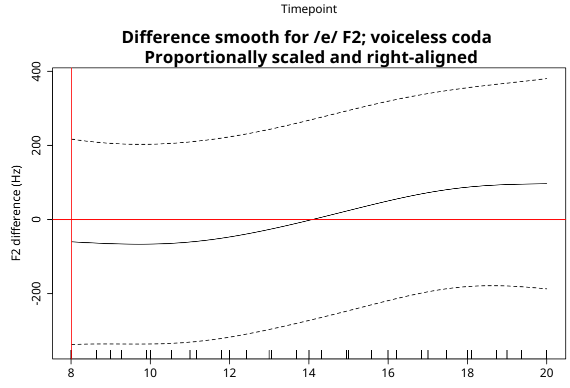

GAMM: /e/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /e/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

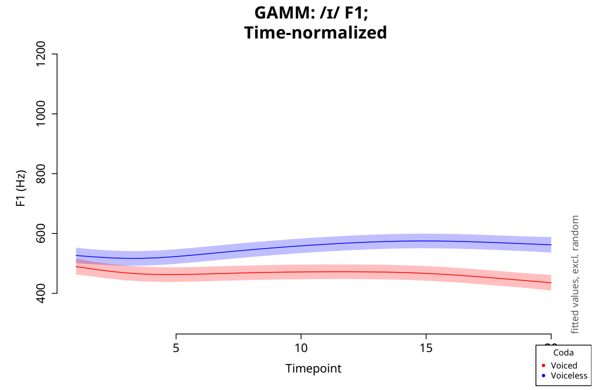

Smooth for /e/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint

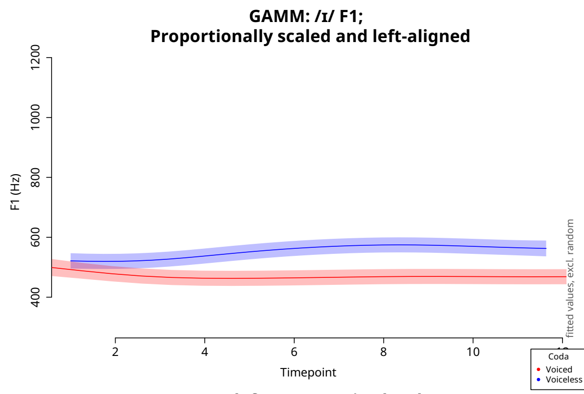

Difference smooth for /e/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint

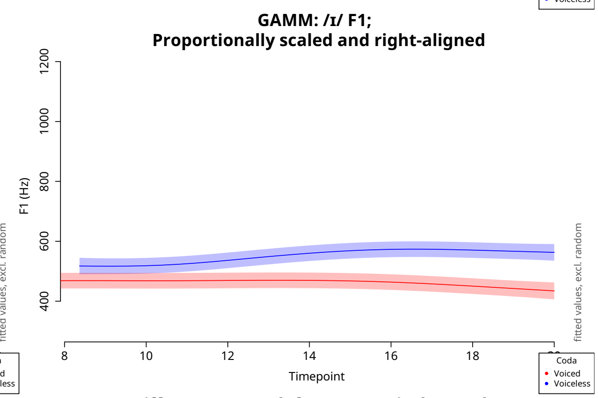

Difference smooth for /e/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

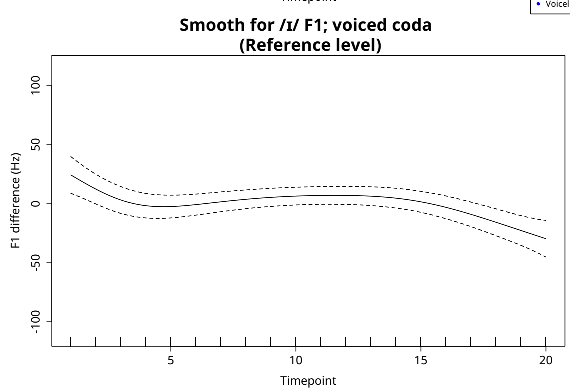

Difference smooth for /e/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

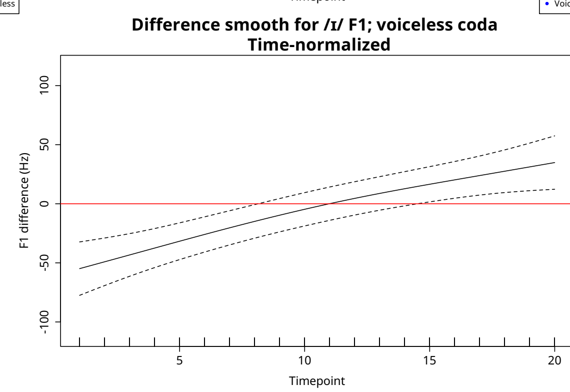
ɪ F1

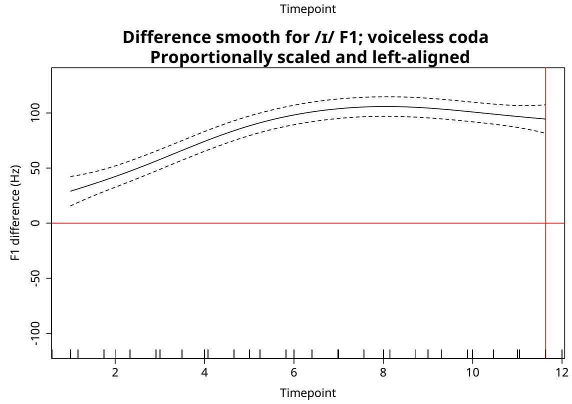

GAMM: /ɪ/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

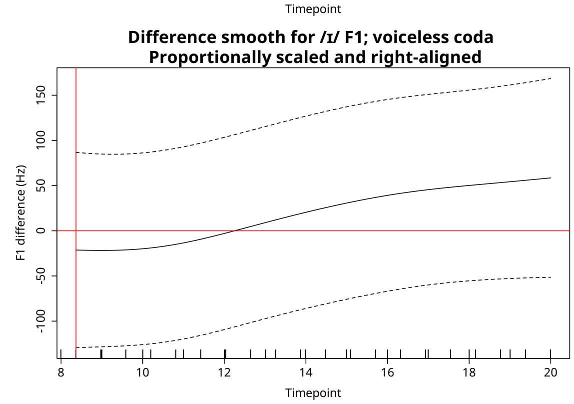

GAMM: /ɪ/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ɪ/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

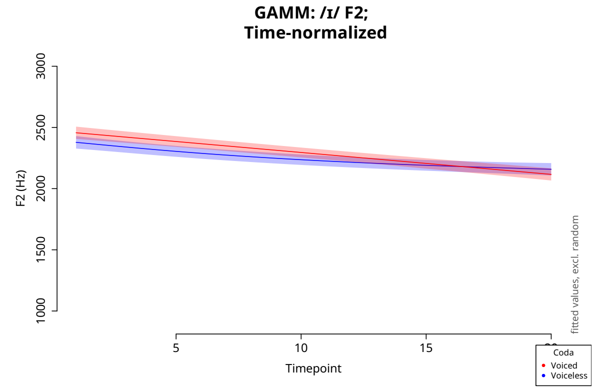

Smooth for /ɪ/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint

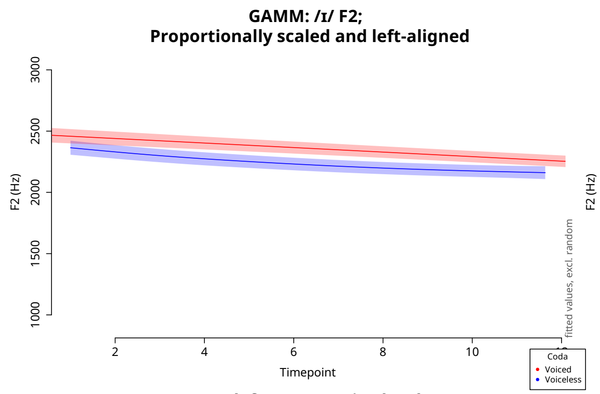

Difference smooth for /ɪ/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint

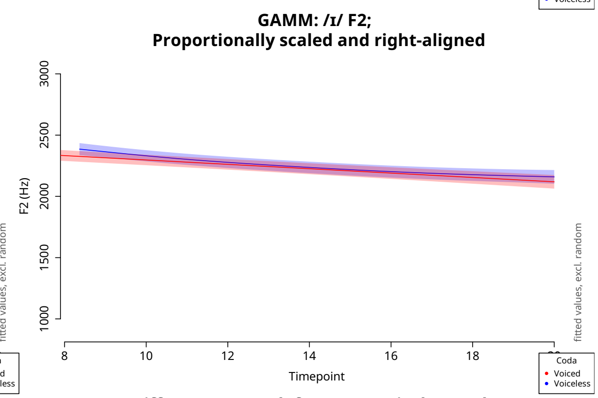

Difference smooth for /ɪ/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

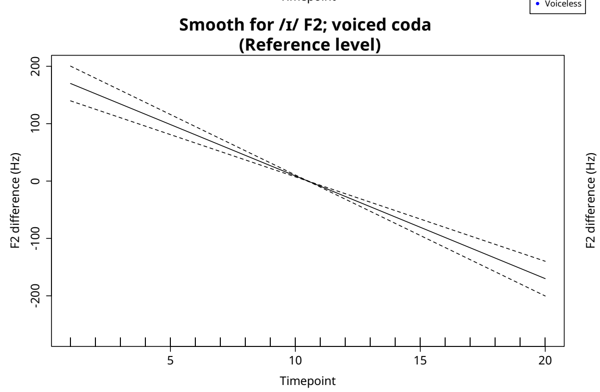

Difference smooth for /ɪ/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

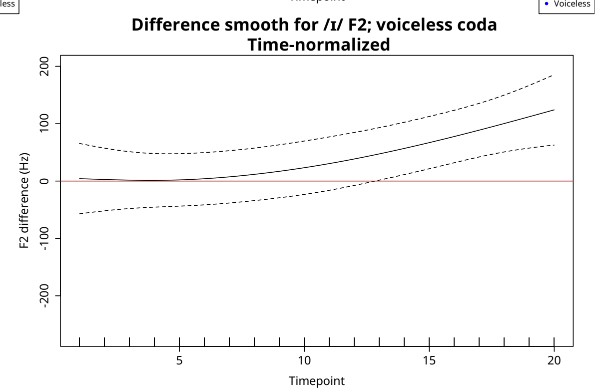
ɪ F2

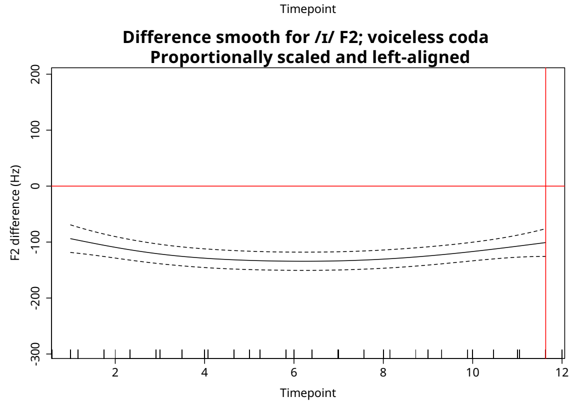

GAMM: /ɪ/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

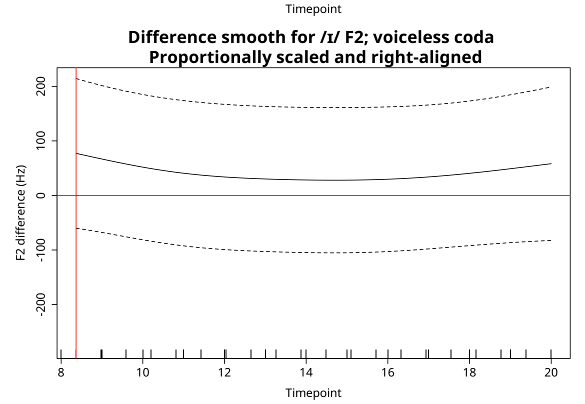

GAMM: /ɪ/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ɪ/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

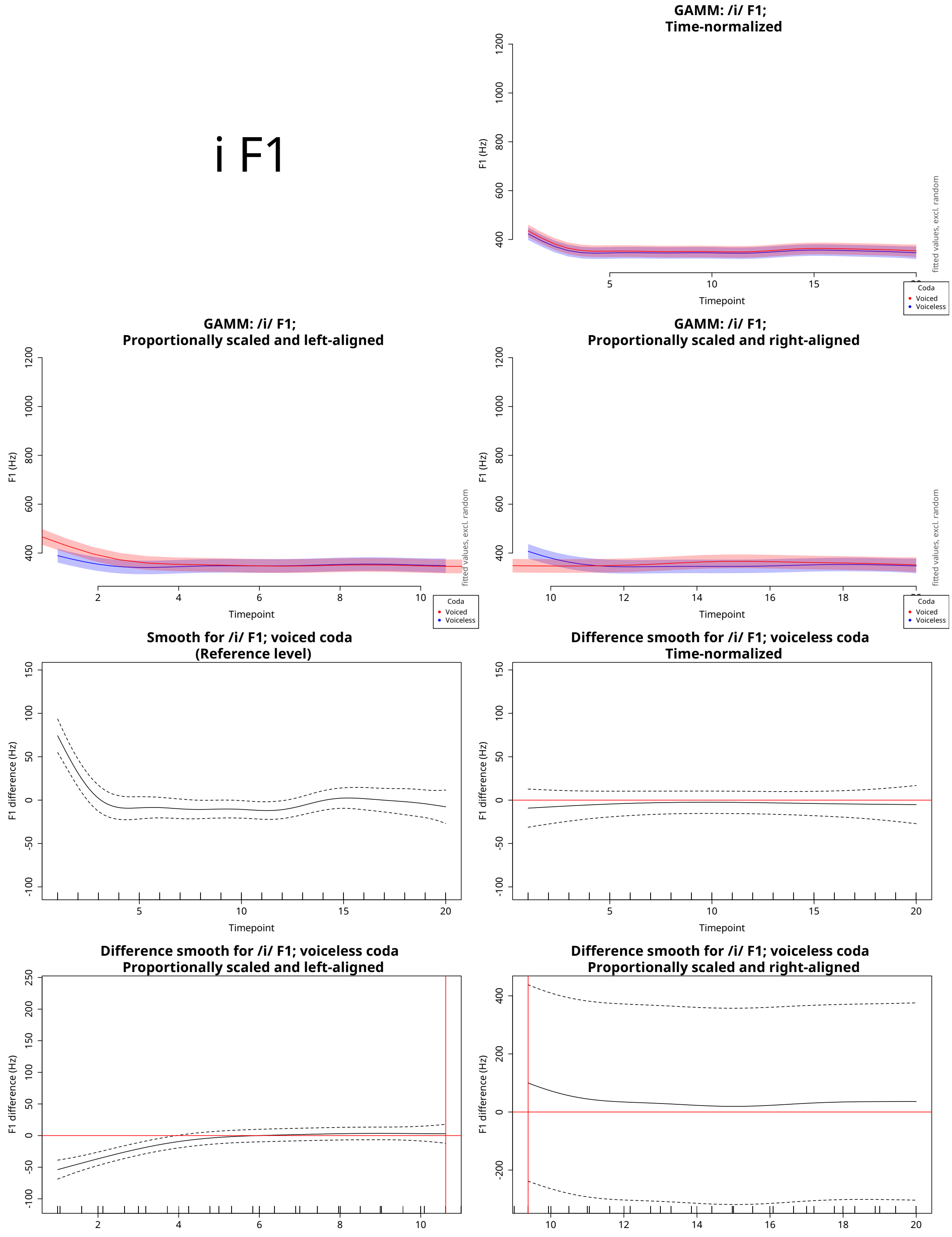

Smooth for /ɪ/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint


Difference smooth for /ɪ/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint


Difference smooth for /ɪ/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

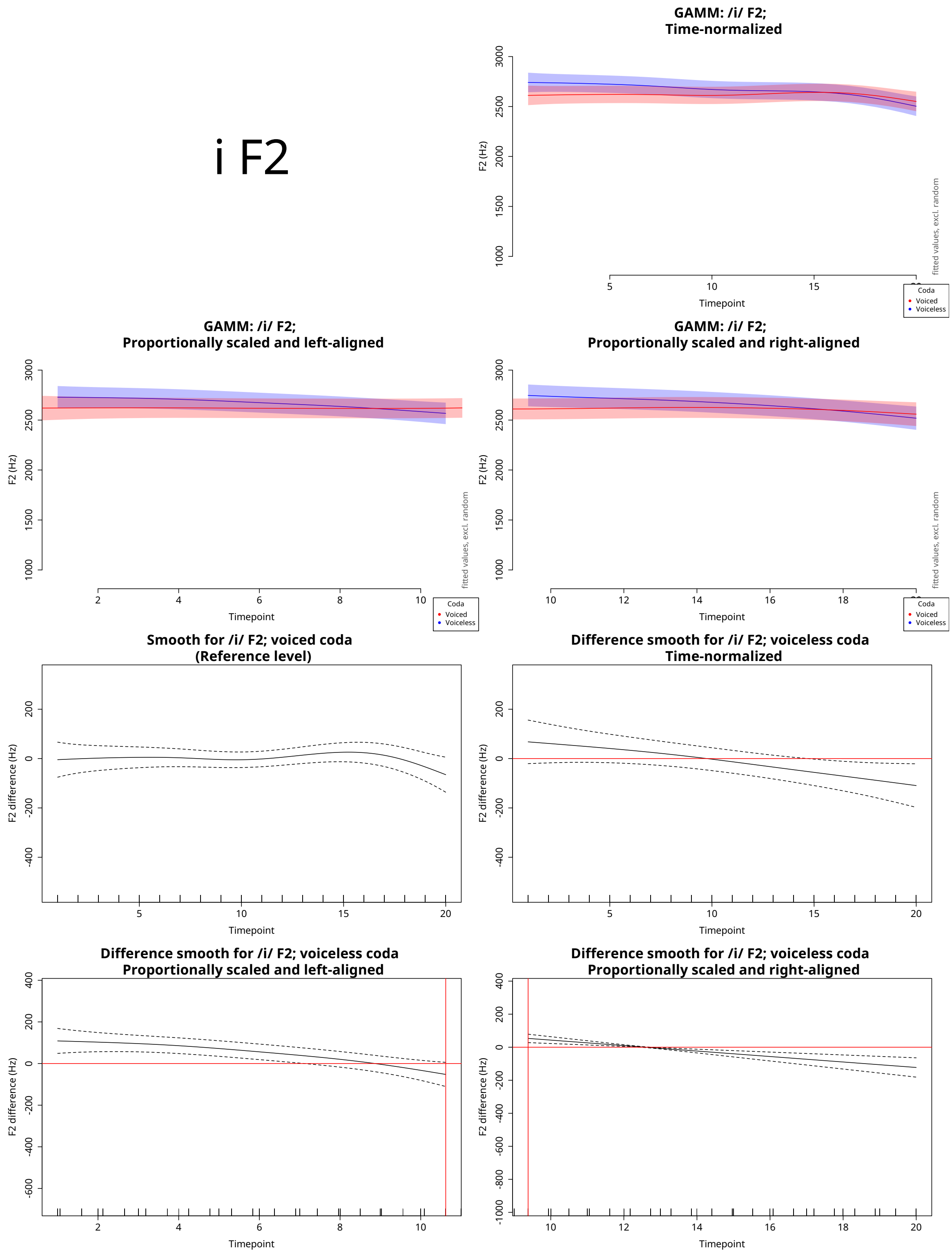

Difference smooth for /ɪ/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

i F1



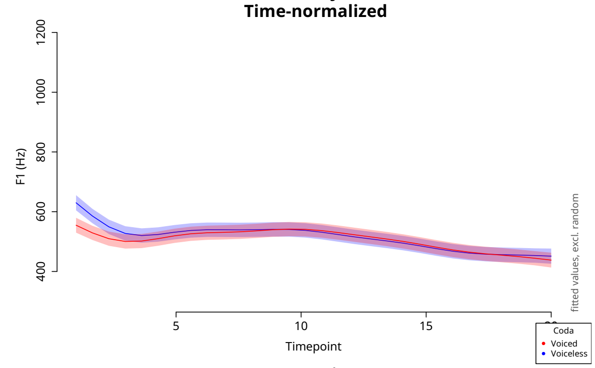
GAMM: /i/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless
GAMM: /i/ F1;
Proportionally scaled and left-aligned
GAMM: /i/ F1;
Proportionally scaled and right-aligned
Smooth for /i/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Difference smooth for /i/ F1; voiceless coda
Time-normalized
Difference smooth for /i/ F1; voiceless coda
Proportionally scaled and left-aligned
Difference smooth for /i/ F1; voiceless coda
Proportionally scaled and right-aligned

i F2

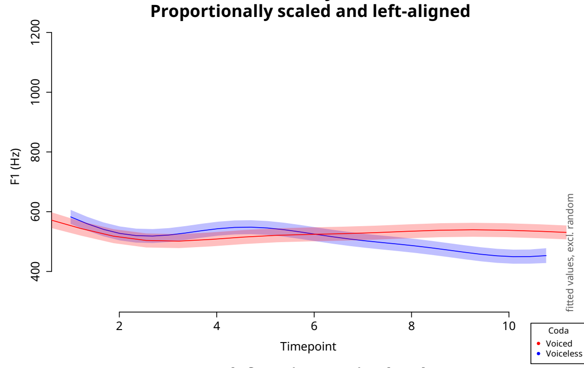

GAMM: /i/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless
GAMM: /i/ F2;
Proportionally scaled and left-aligned
GAMM: /i/ F2;
Proportionally scaled and right-aligned
Smooth for /i/ F2; voiced coda
(Reference level)
Difference smooth for /i/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Difference smooth for /i/ F2; voiceless coda
Proportionally scaled and left-aligned
Difference smooth for /i/ F2; voiceless coda
Proportionally scaled and right-aligned

ɔj F1

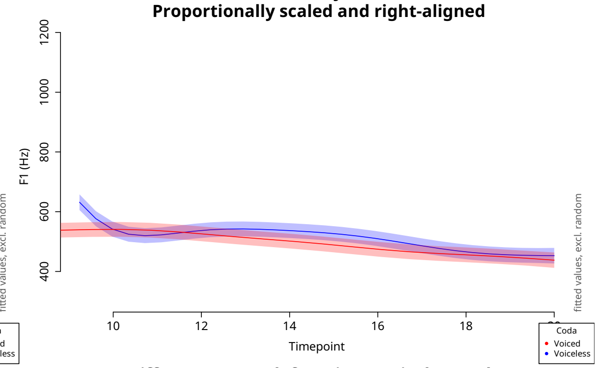
GAMM: /ɔj/ F1;
Time-normalized



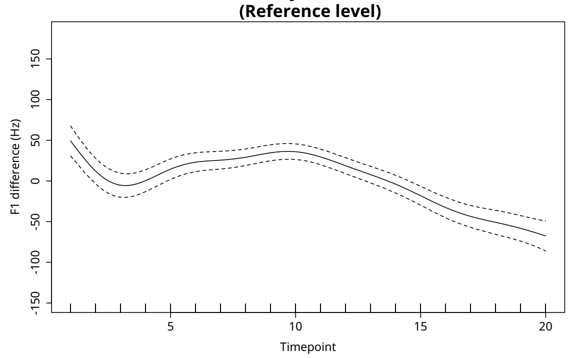
GAMM: /ɔj/ F1;
Proportionally scaled and left-aligned



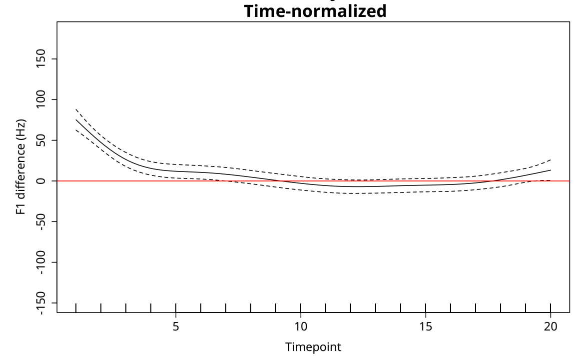
GAMM: /ɔj/ F1;
Proportionally scaled and right-aligned



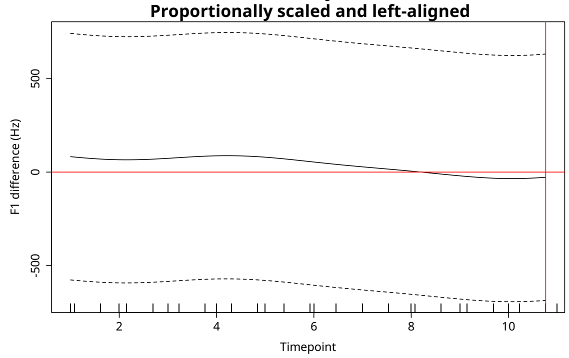
Smooth for /ɔj/ F1; voiced coda
(Reference level)



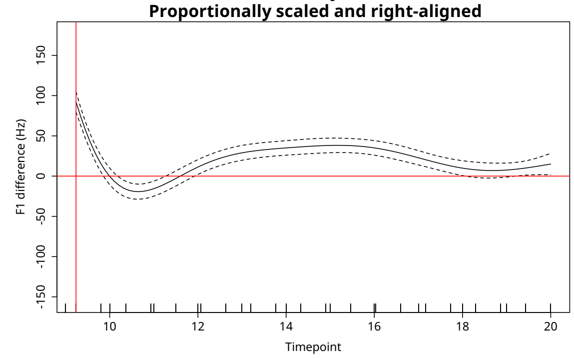
Difference smooth for /ɔj/ F1; voiceless coda
Time-normalized



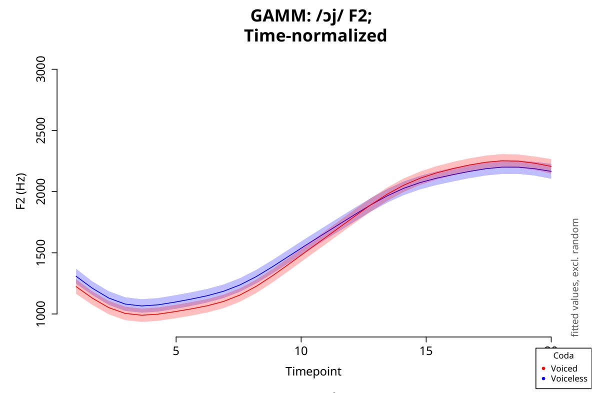
Difference smooth for /ɔj/ F1; voiceless coda
Proportionally scaled and left-aligned

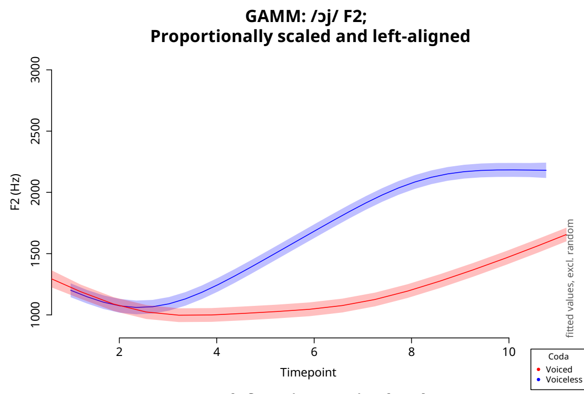


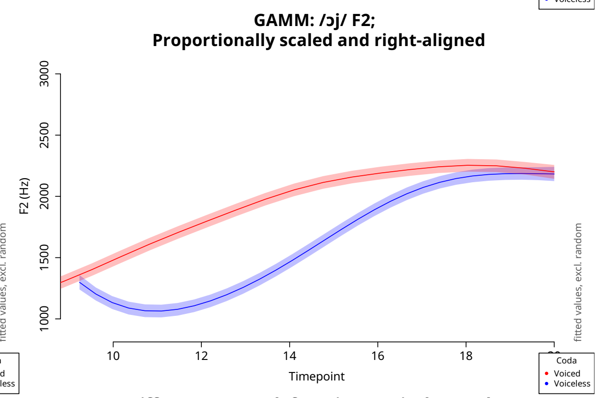
Difference smooth for /ɔj/ F1; voiceless coda
Proportionally scaled and right-aligned

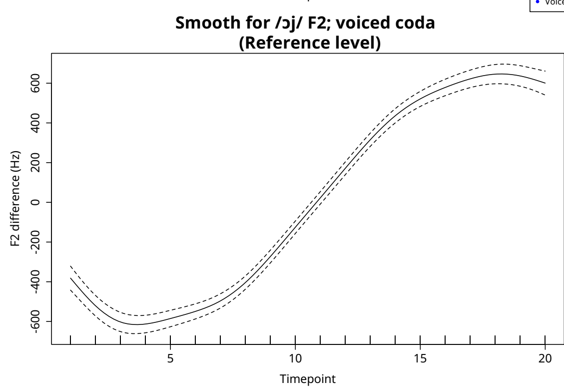


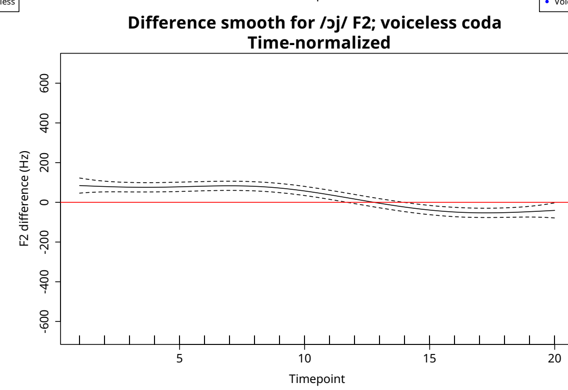
ɔj F2

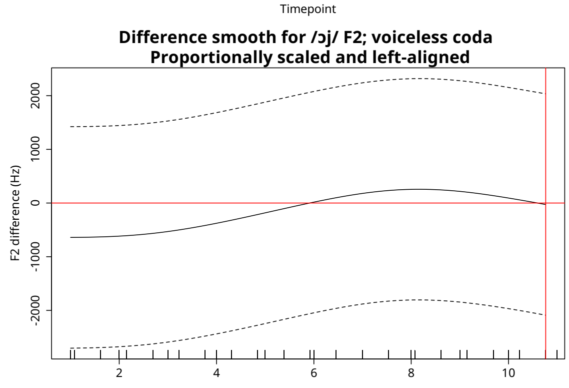

GAMM: /ɔj/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

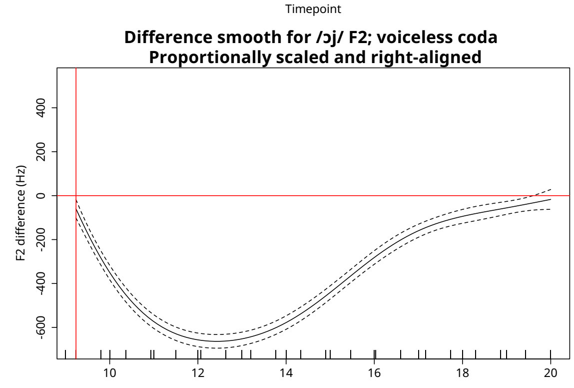

GAMM: /ɔj/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ɔj/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


Smooth for /ɔj/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint

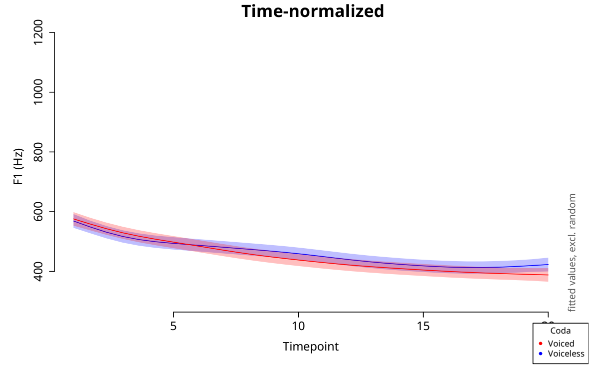

Difference smooth for /ɔj/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint


Difference smooth for /ɔj/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

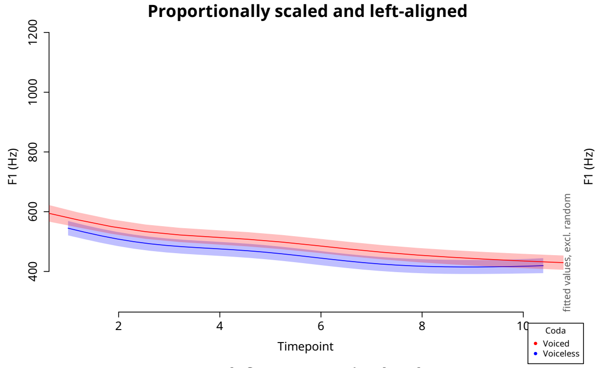

Difference smooth for /ɔj/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

o F1

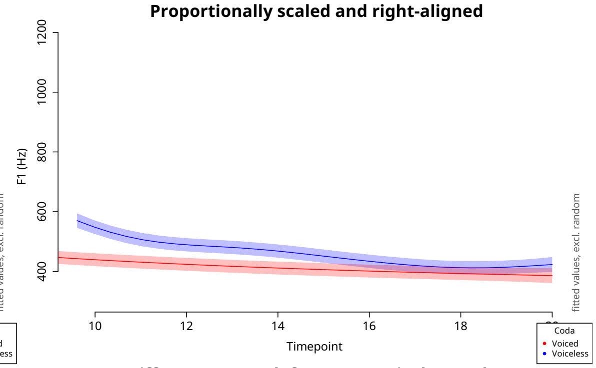
GAMM: /o/ F1;
Time-normalized



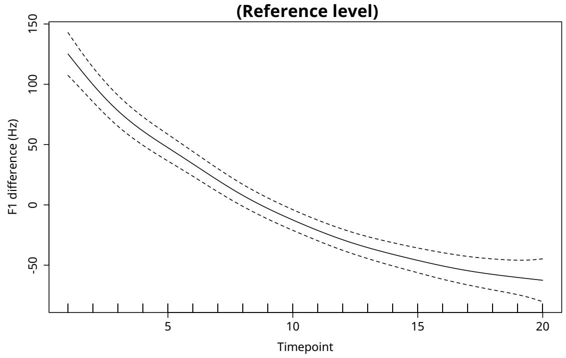
GAMM: /o/ F1;
Proportionally scaled and left-aligned



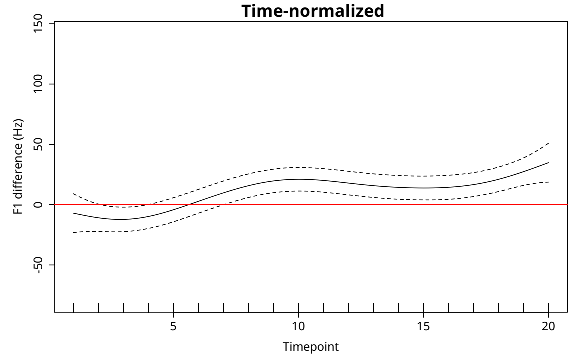
GAMM: /o/ F1;
Proportionally scaled and right-aligned



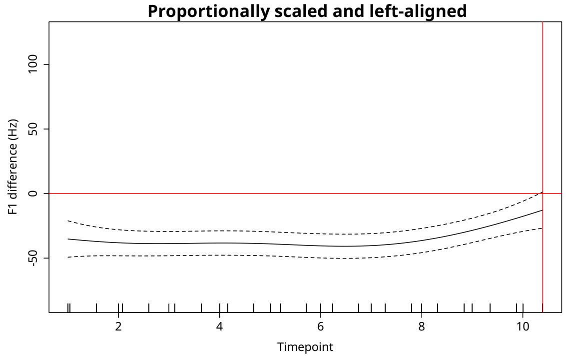
Smooth for /o/ F1; voiced coda
(Reference level)



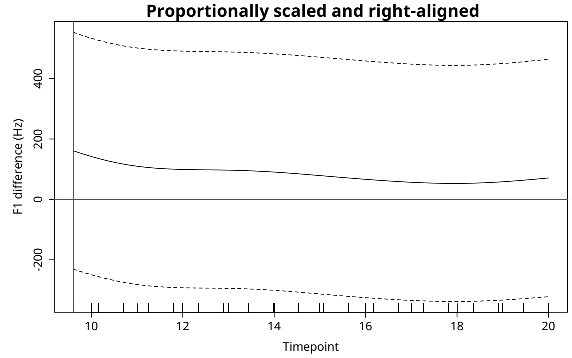
Difference smooth for /o/ F1; voiceless coda
Time-normalized



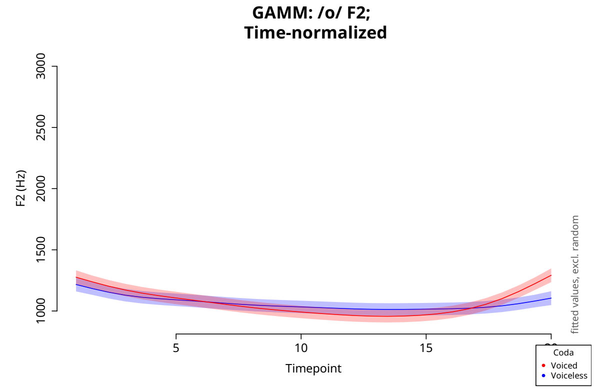
Difference smooth for /o/ F1; voiceless coda
Proportionally scaled and left-aligned

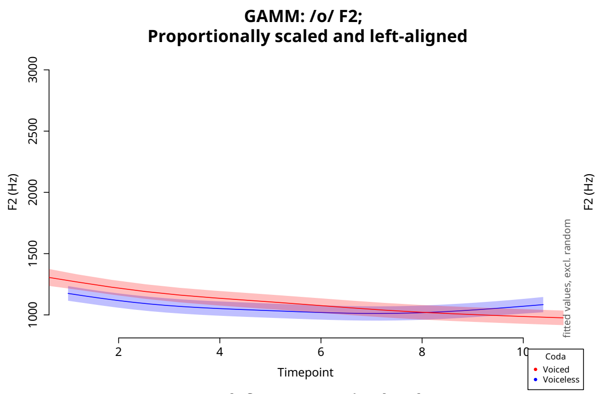


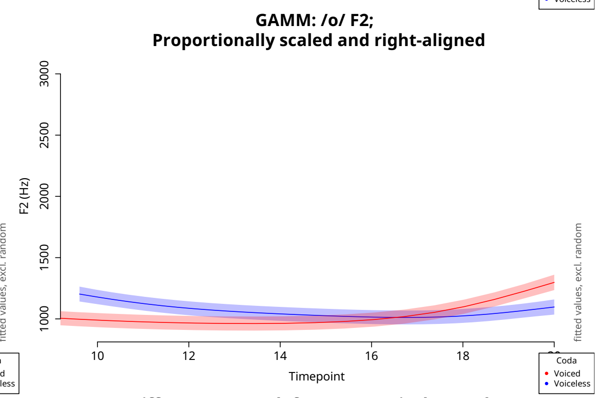
Difference smooth for /o/ F1; voiceless coda
Proportionally scaled and right-aligned

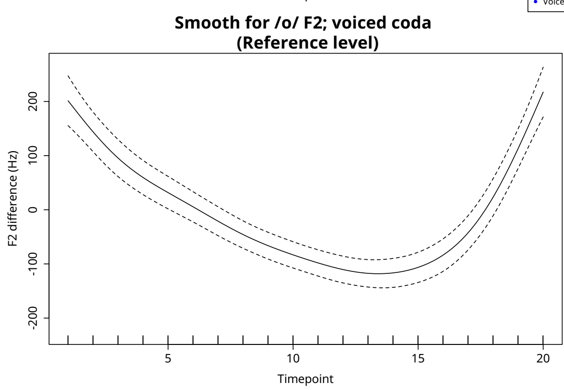


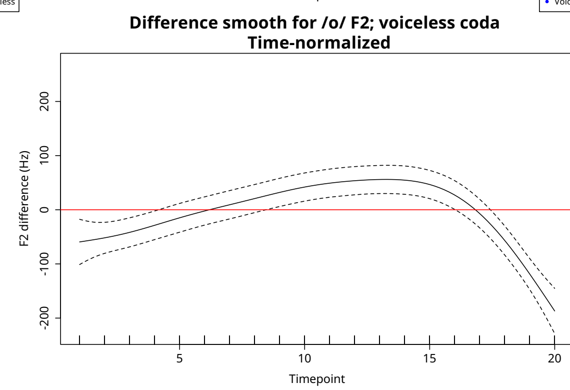
o F2

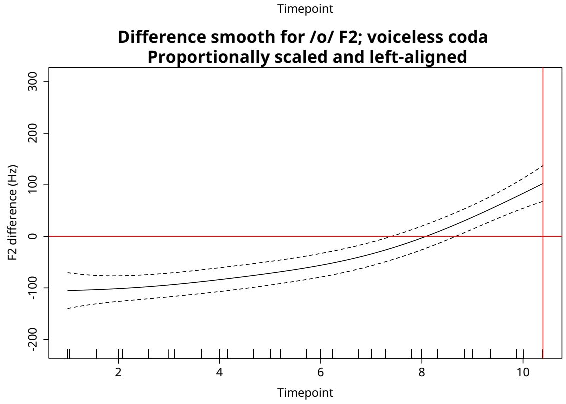

GAMM: /o/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

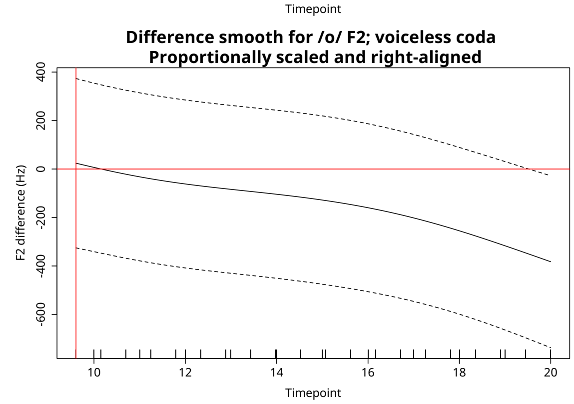

GAMM: /o/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /o/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

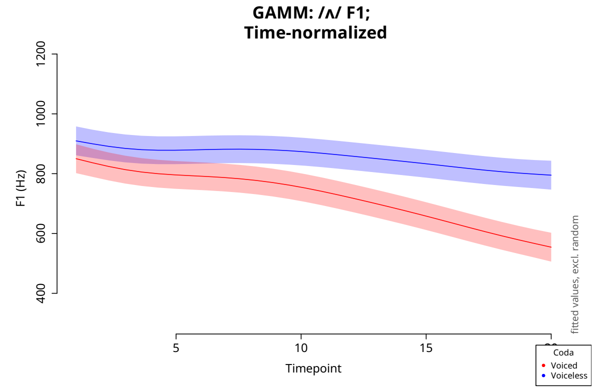

Smooth for /o/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint

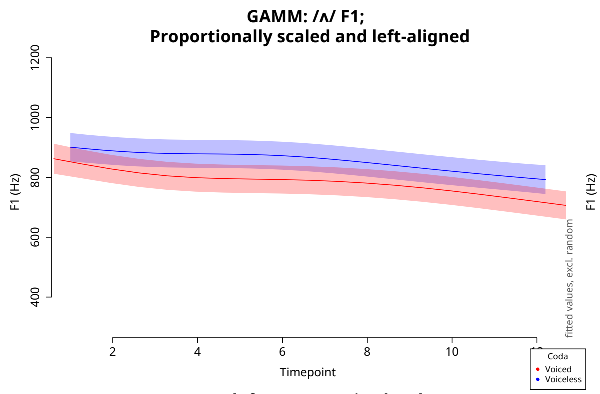

Difference smooth for /o/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint

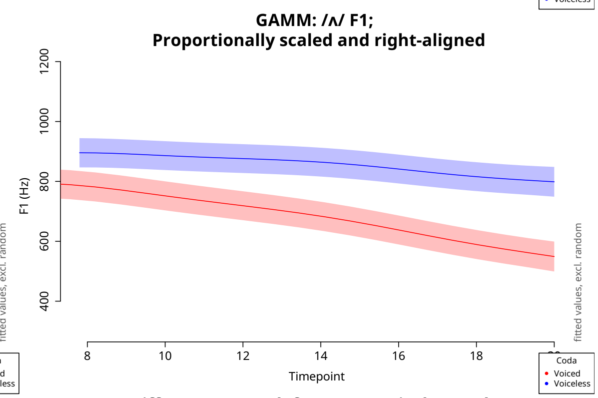

Difference smooth for /o/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

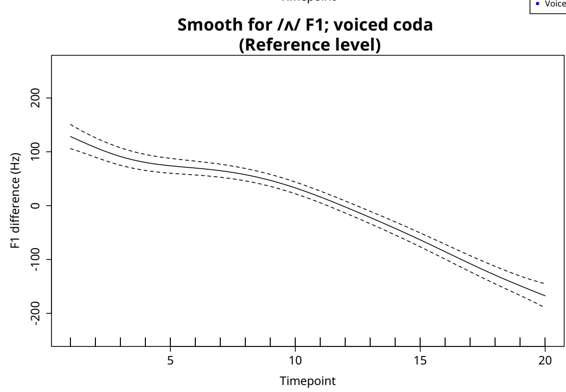

Difference smooth for /o/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

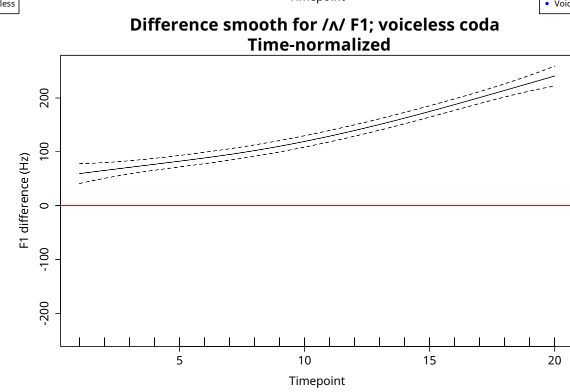
ʌ F1

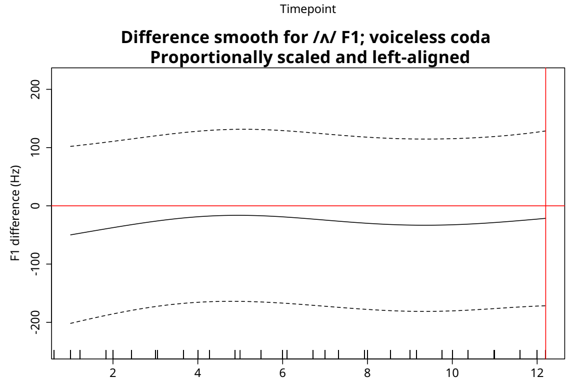

GAMM: /ʌ/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

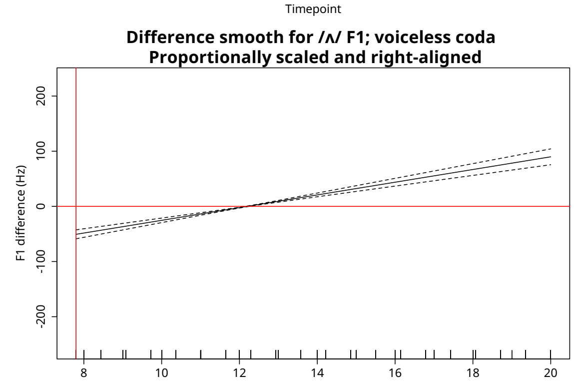

GAMM: /ʌ/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ʌ/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

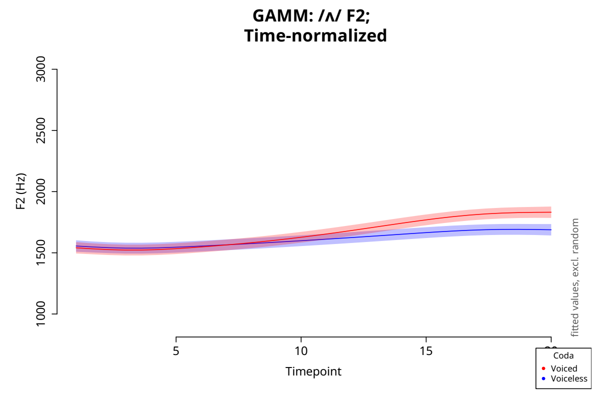

Smooth for /ʌ/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint

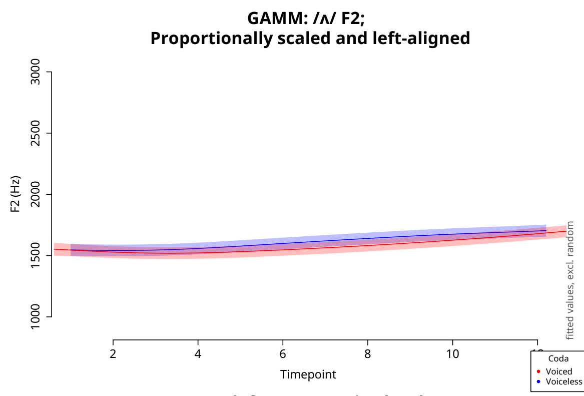

Difference smooth for /ʌ/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint

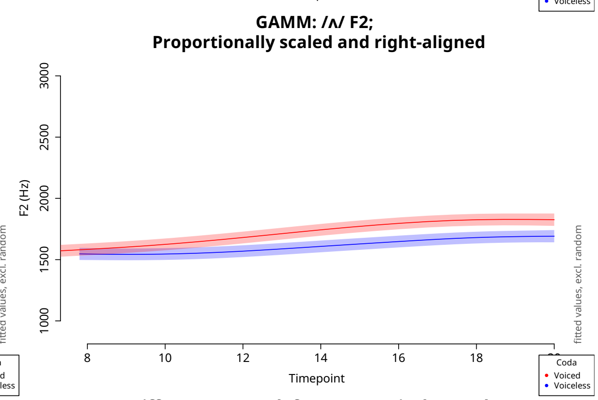

Difference smooth for /ʌ/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

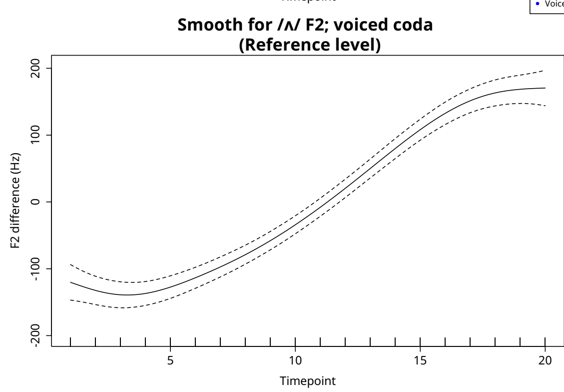

Difference smooth for /ʌ/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

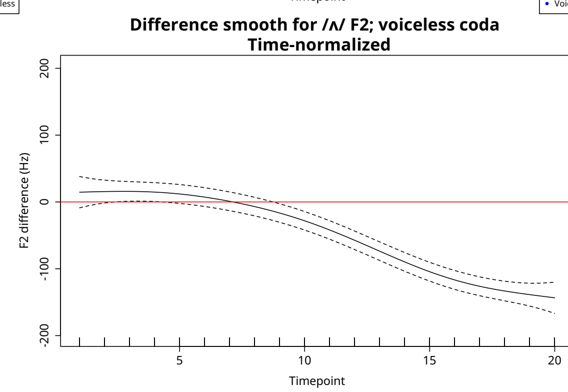
ʌ F2

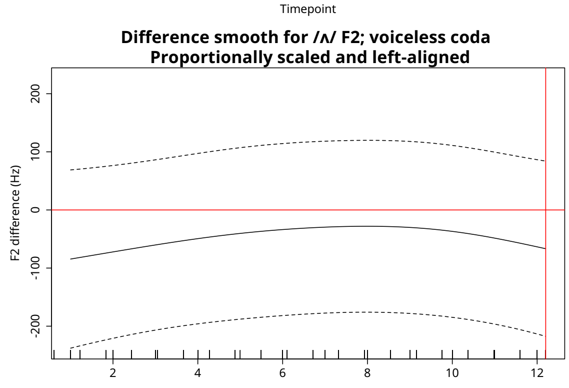

GAMM: /ʌ/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

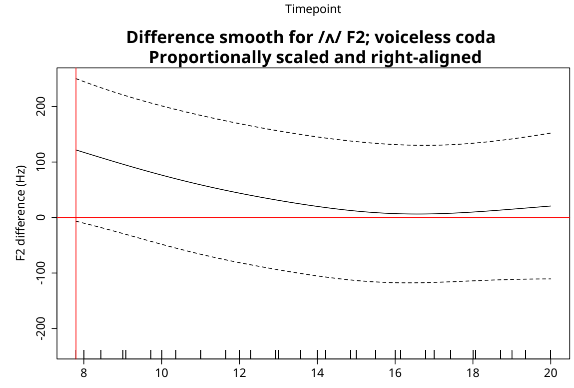

GAMM: /ʌ/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ʌ/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

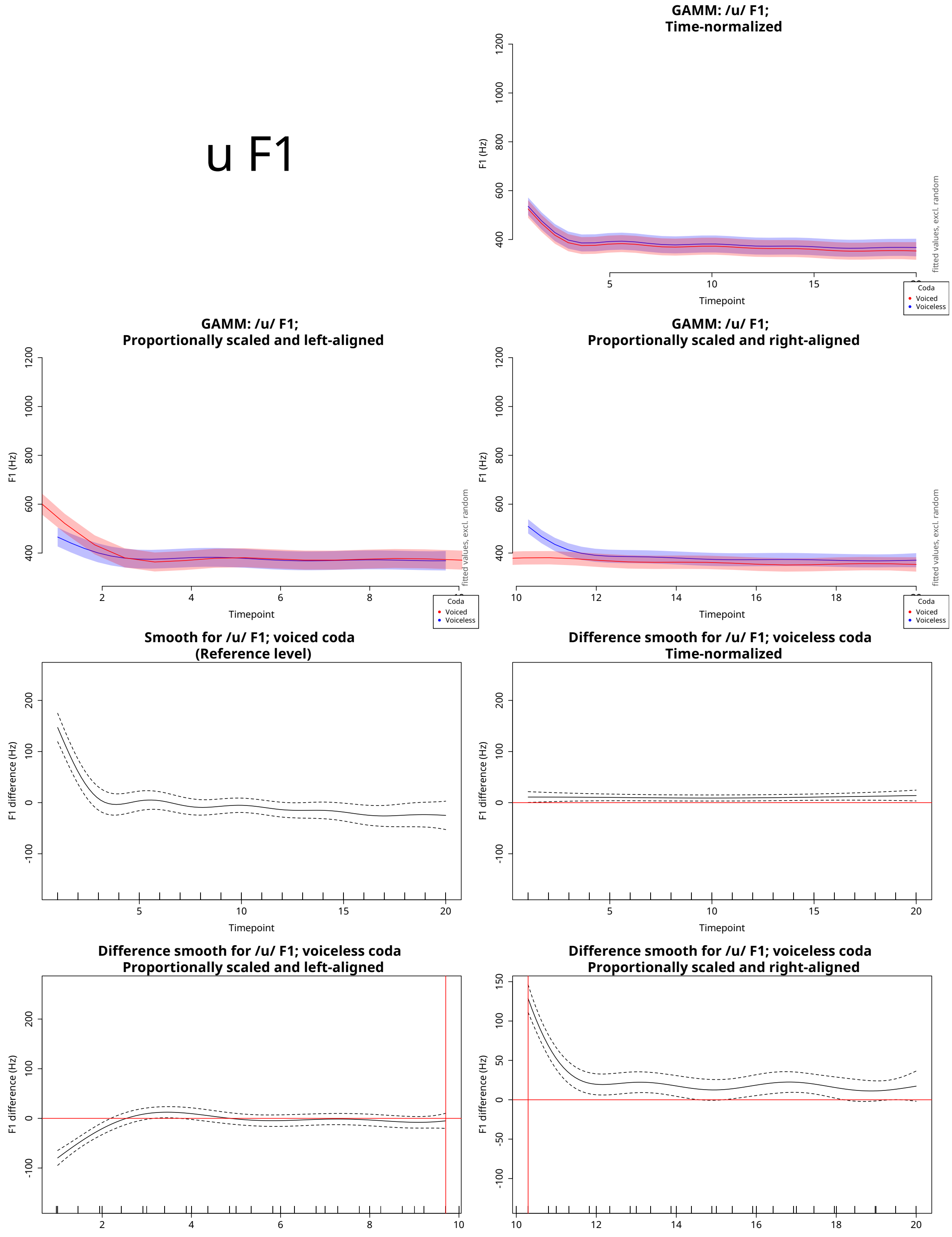

Smooth for /ʌ/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint


Difference smooth for /ʌ/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint


Difference smooth for /ʌ/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint


Difference smooth for /ʌ/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

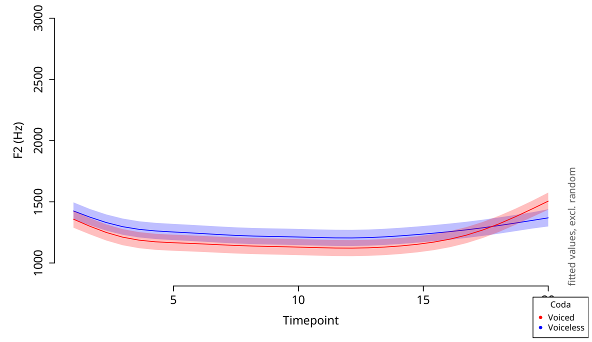
u F1



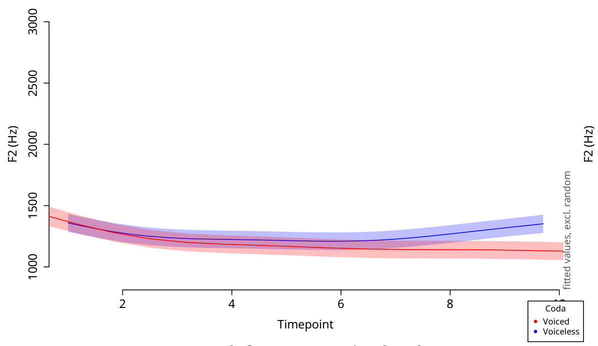
GAMM: /u/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless
GAMM: /u/ F1;
Proportionally scaled and left-aligned
GAMM: /u/ F1;
Proportionally scaled and right-aligned
Smooth for /u/ F1; voiced coda
(Reference level)
Difference smooth for /u/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Difference smooth for /u/ F1; voiceless coda
Proportionally scaled and left-aligned
Difference smooth for /u/ F1; voiceless coda
Proportionally scaled and right-aligned

u F2

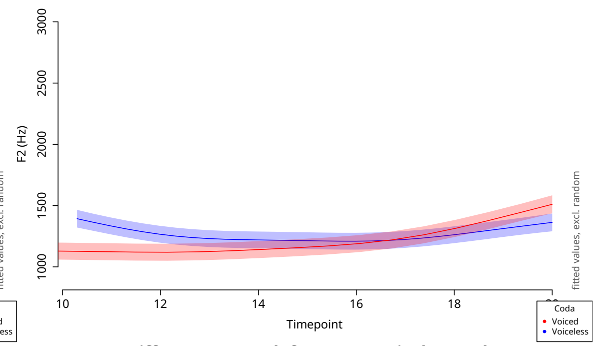
GAMM: /u/ F2;
Time-normalized



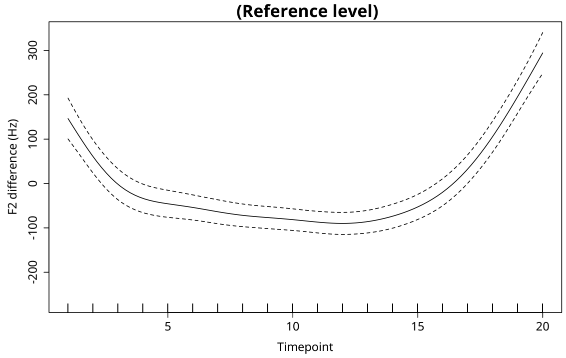
GAMM: /u/ F2;
Proportionally scaled and left-aligned



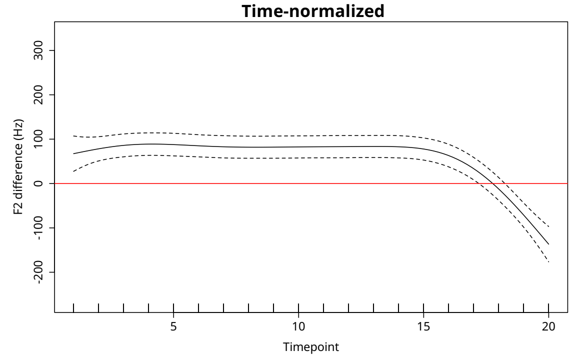
GAMM: /u/ F2;
Proportionally scaled and right-aligned



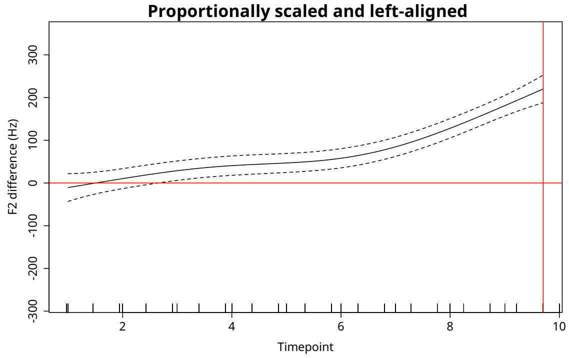
Smooth for /u/ F2; voiced coda
(Reference level)



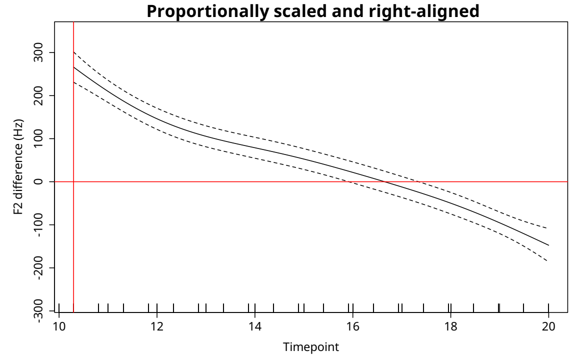
Difference smooth for /u/ F2; voiceless coda
Time-normalized



Difference smooth for /u/ F2; voiceless coda
Proportionally scaled and left-aligned



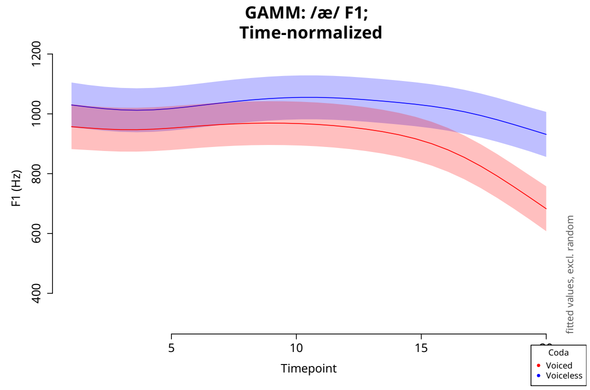
Difference smooth for /u/ F2; voiceless coda
Proportionally scaled and right-aligned

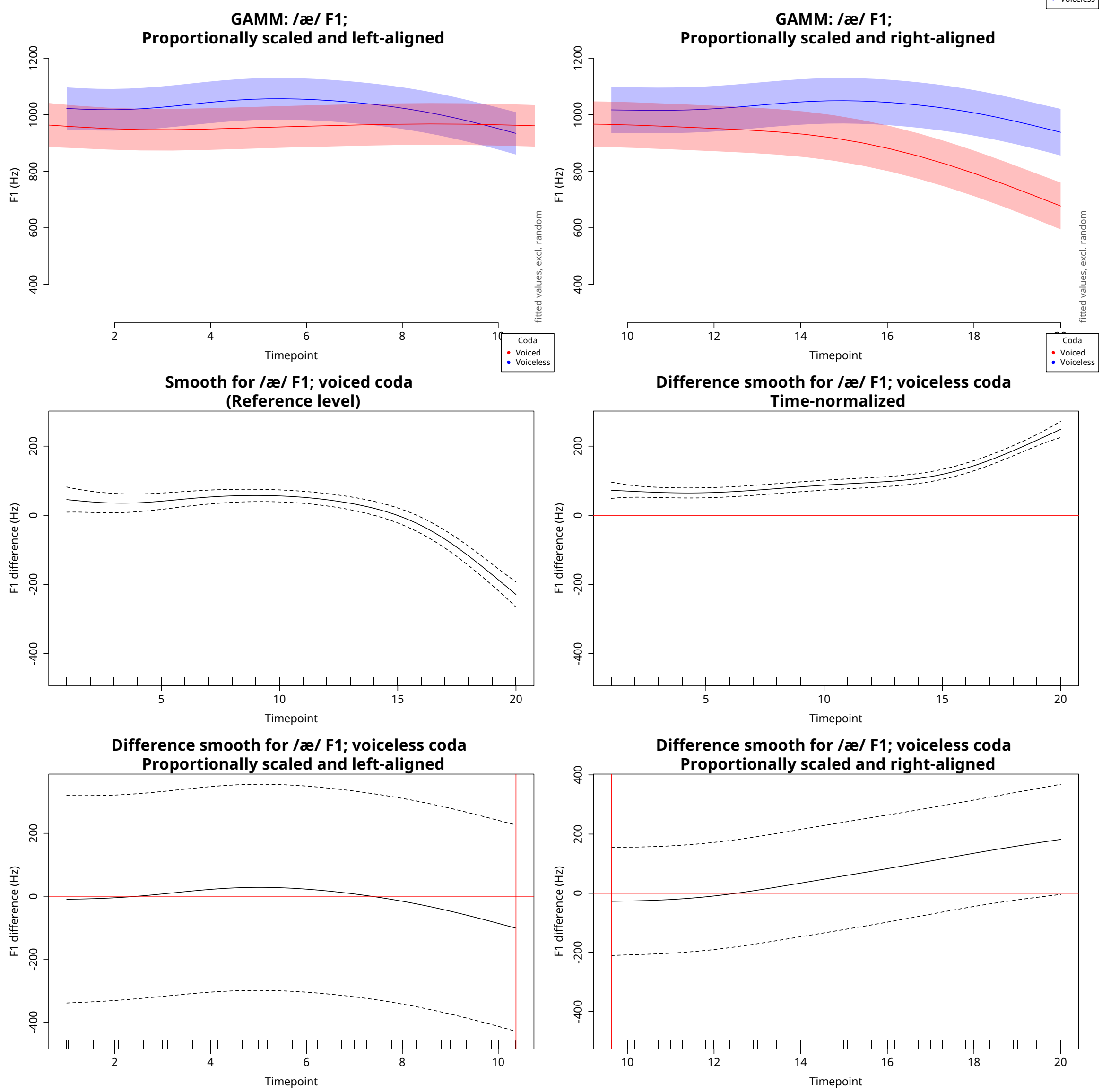


Appendix F GAMMs comparisons, Denver

This appendix contains the full set of GAMMs comparisons for each durational scaling /alignment model (see §4.4.3) across all vowels in the Denver dataset. The top set of three plots (in colour) display smooths with confidence intervals for each coda voicing condition for each model; the lower four plots are difference smooths, with the reference level (voiced coda) at top right, followed by the time-normalized model at top left, and the proportionally-scaled duration models, left- and right-alignment, in the lower left and right plots, respectively.

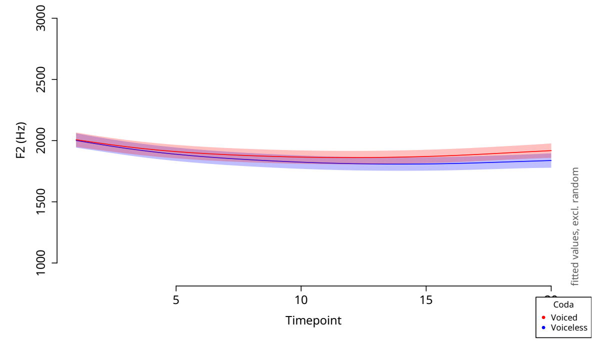
æ F1


GAMM: /æ/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

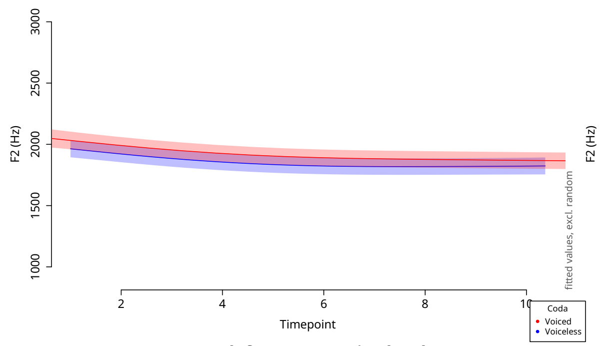

GAMM: /æ/ F1;
Proportionally scaled and left-aligned
GAMM: /æ/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless
Smooth for /æ/ F1; voiced coda
(Reference level)
Difference smooth for /æ/ F1; voiceless coda
Time-normalized
Difference smooth for /æ/ F1; voiceless coda
Proportionally scaled and left-aligned
Difference smooth for /æ/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

æ F2

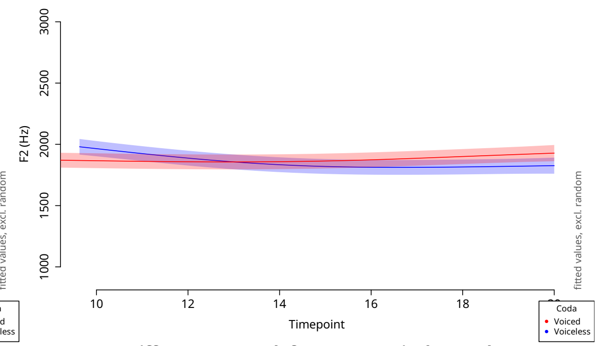
GAMM: /æ/ F2;
Time-normalized



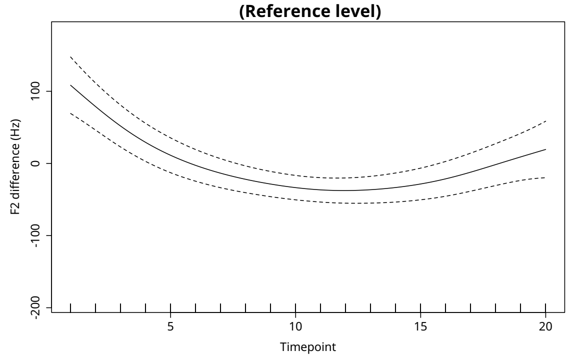
GAMM: /æ/ F2;
Proportionally scaled and left-aligned



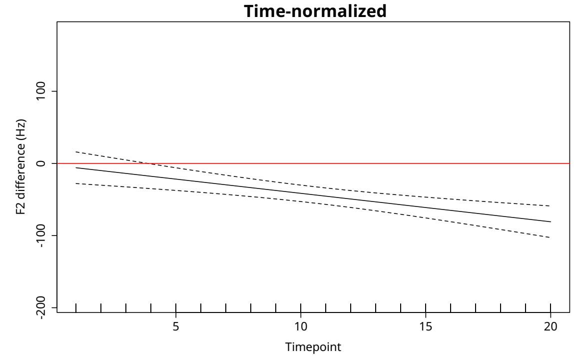
GAMM: /æ/ F2;
Proportionally scaled and right-aligned



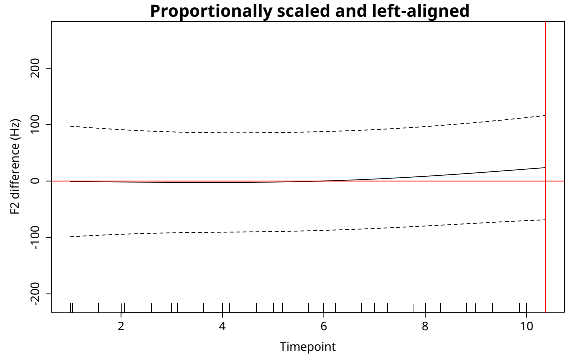
Smooth for /æ/ F2; voiced coda
(Reference level)



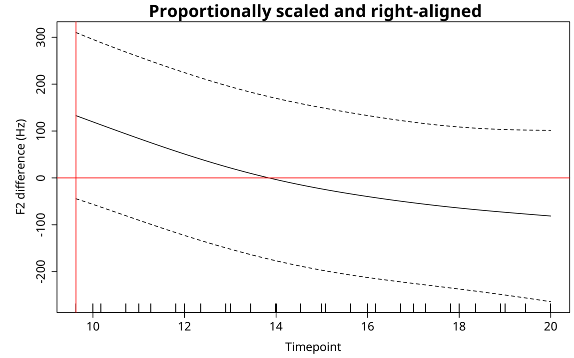
Difference smooth for /æ/ F2; voiceless coda
Time-normalized



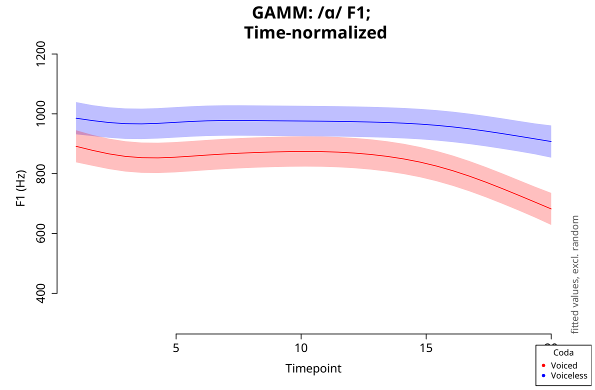
Difference smooth for /æ/ F2; voiceless coda
Proportionally scaled and left-aligned

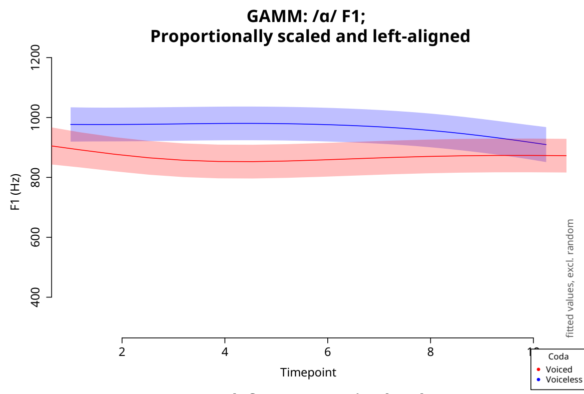


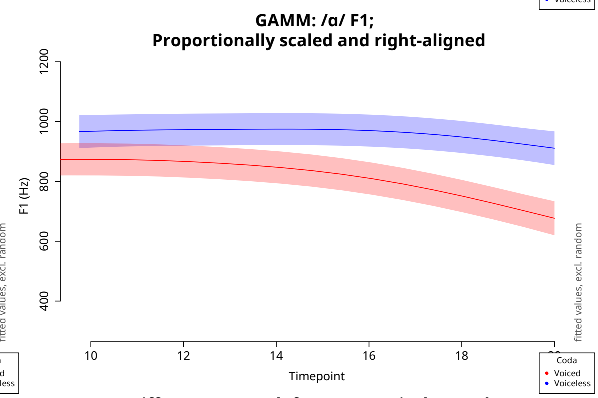
Difference smooth for /æ/ F2; voiceless coda
Proportionally scaled and right-aligned

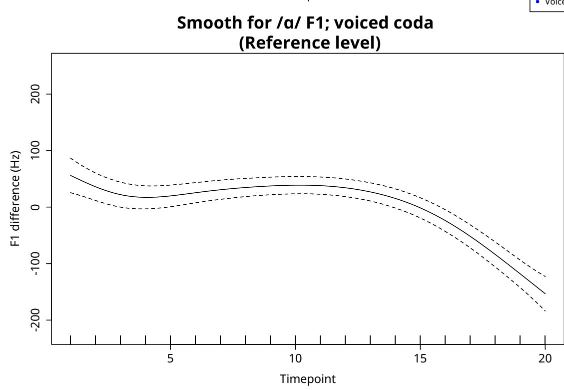


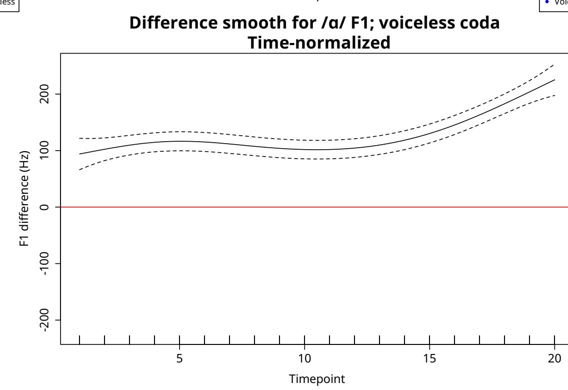
ɑ F1

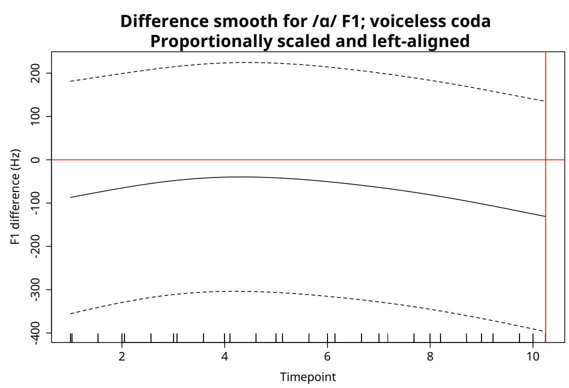

GAMM: /ɑ/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

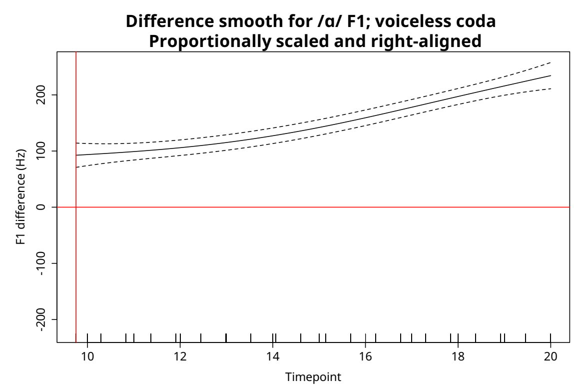

GAMM: /ɑ/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ɑ/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

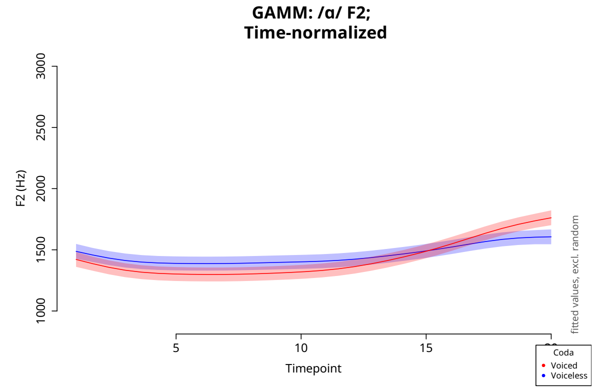

Smooth for /ɑ/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint

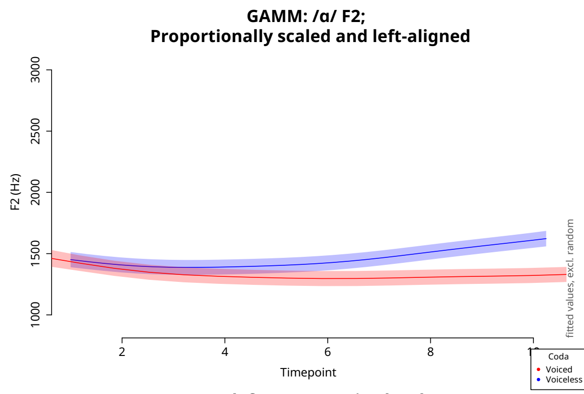

Difference smooth for /ɑ/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint

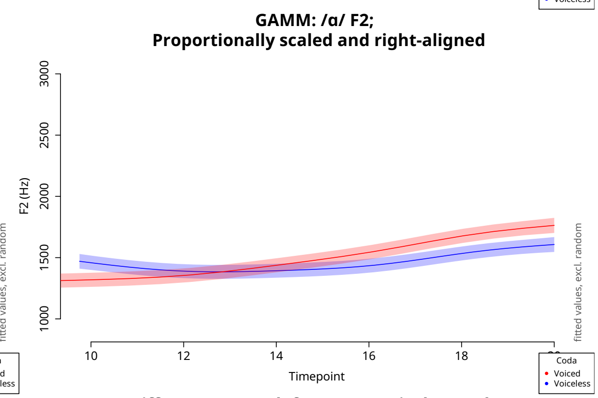

Difference smooth for /ɑ/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

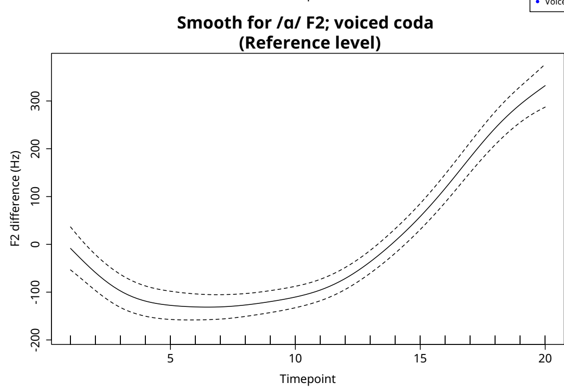

Difference smooth for /ɑ/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

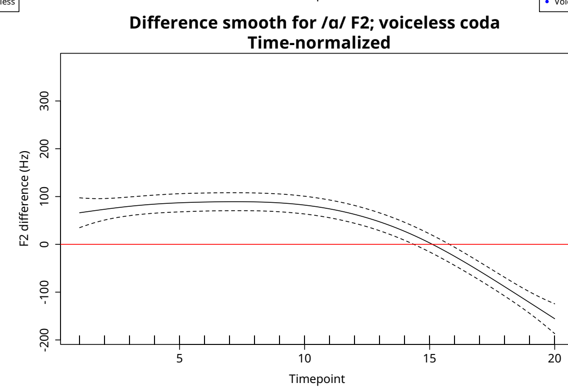
ɑ F2

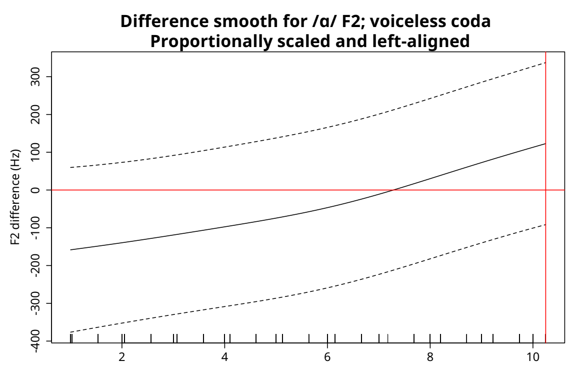

GAMM: /ɑ/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

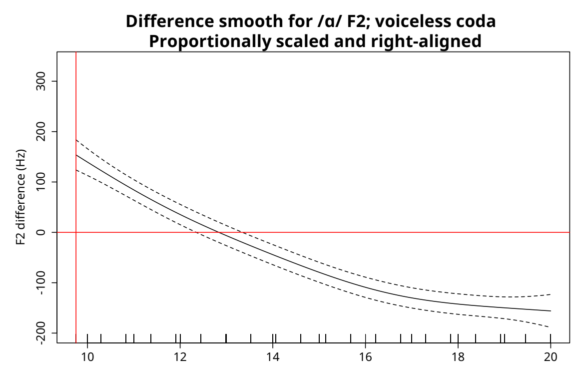

GAMM: /ɑ/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ɑ/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

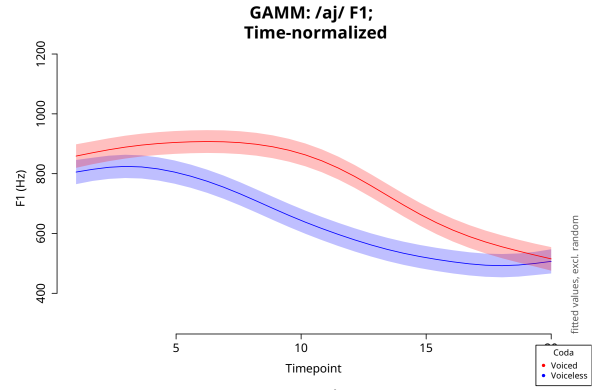

Smooth for /ɑ/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint

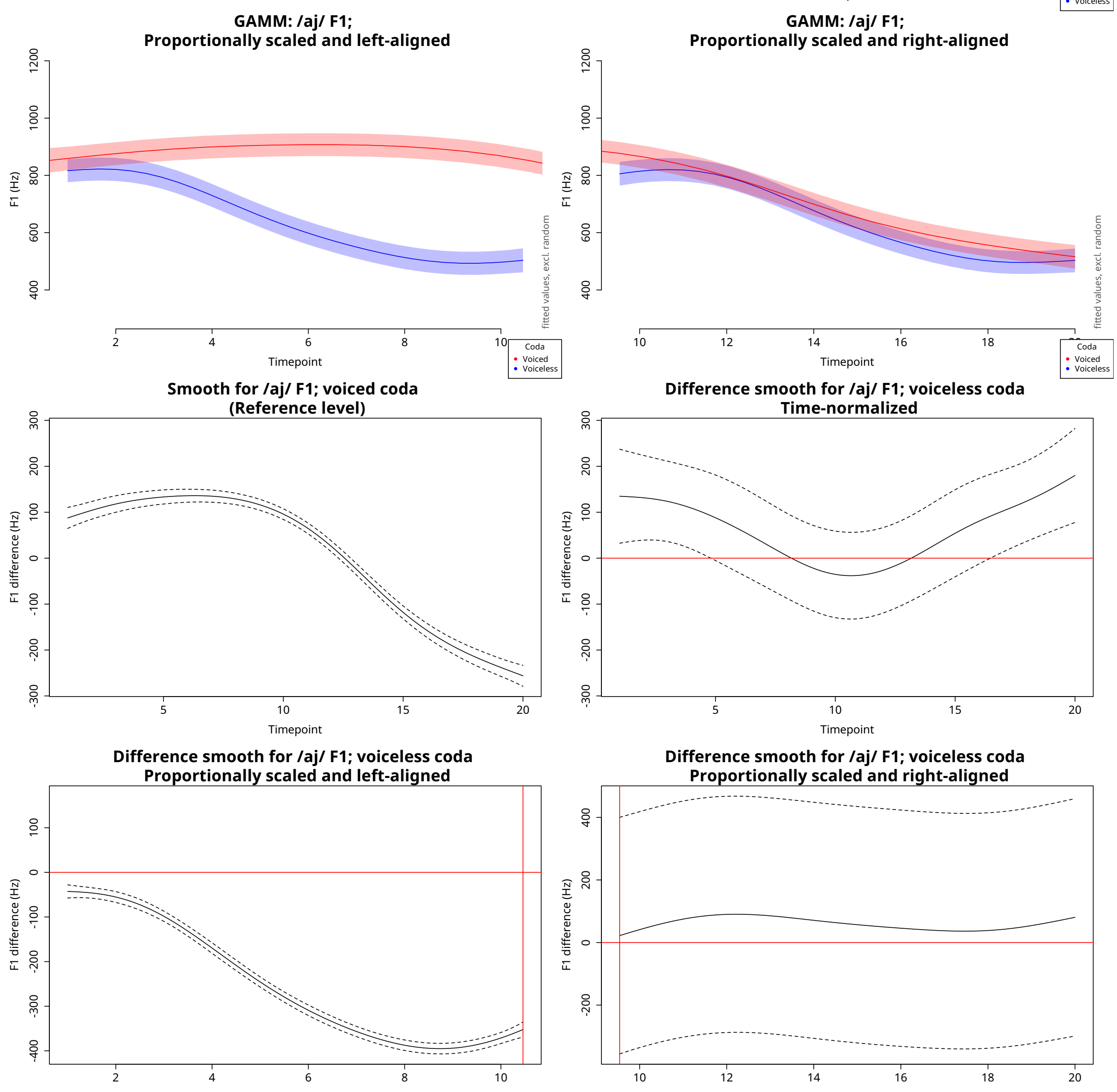

Difference smooth for /ɑ/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint


Difference smooth for /ɑ/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint


Difference smooth for /ɑ/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

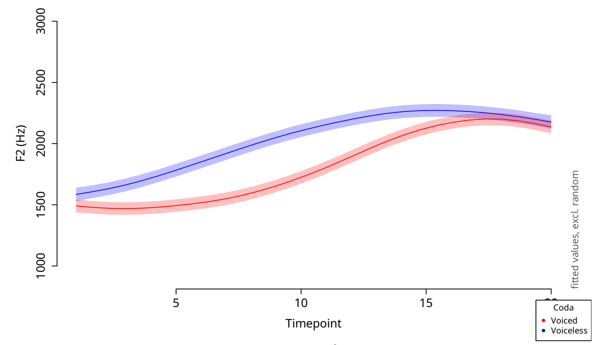
aj F1


GAMM: /aj/ F1;
Time-normalized
F1 (Hz)
400
600
800
1000
1200
5
10
15
20
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

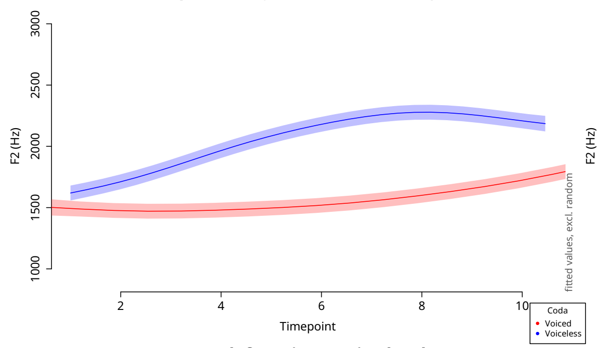

GAMM: /aj/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
400
600
800
1000
1200
2
4
6
8
10
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless
GAMM: /aj/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
10
12
14
16
18
20
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless
Smooth for /aj/ F1; voiced coda
(Reference level)
F1 difference (Hz)
-300
-200
-100
0
100
200
300
5
10
15
20
Timepoint
Difference smooth for /aj/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint
Difference smooth for /aj/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
-400
-300
-200
-100
0
100
2
4
6
8
10
Timepoint
Difference smooth for /aj/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
-200
0
200
400
10
12
14
16
18
20
Timepoint

aj F2

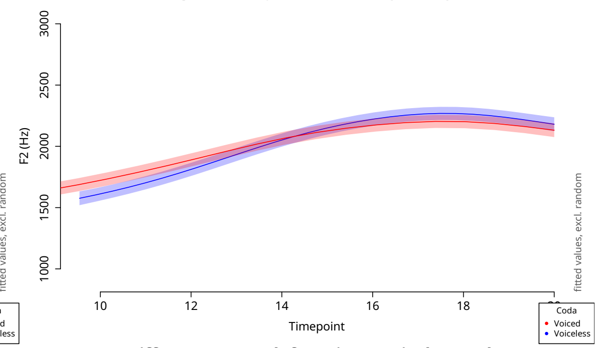
GAMM: /aj/ F2; Time-normalized



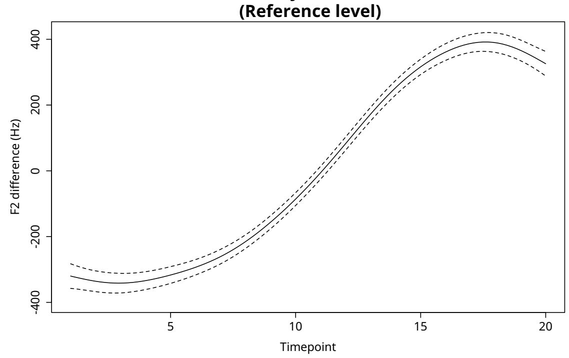
GAMM: /aj/ F2; Proportionally scaled and left-aligned



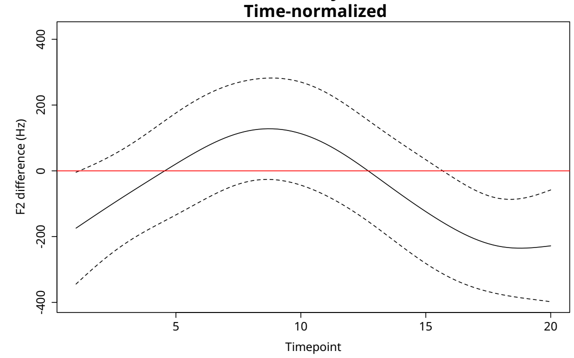
GAMM: /aj/ F2; Proportionally scaled and right-aligned



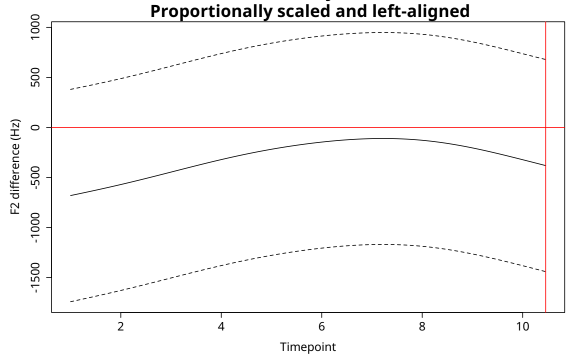
Smooth for /aj/ F2; voiced coda (Reference level)



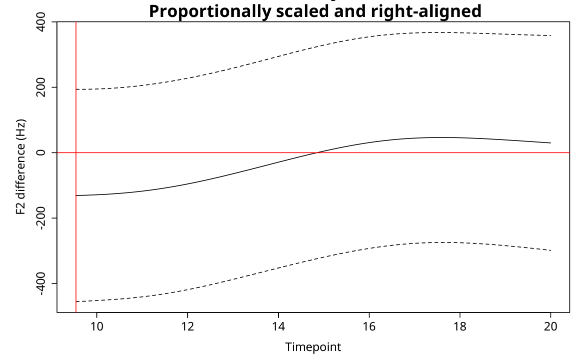
Difference smooth for /aj/ F2; voiceless coda Time-normalized



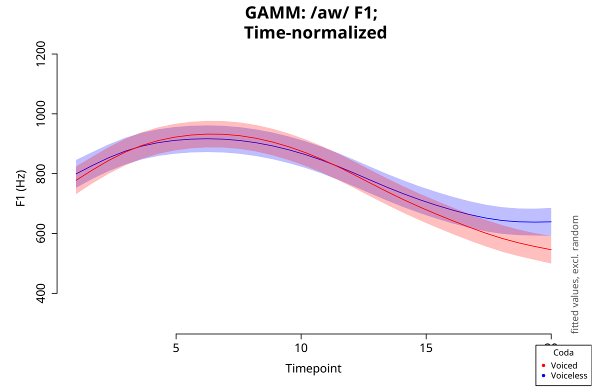
Difference smooth for /aj/ F2; voiceless coda Proportionally scaled and left-aligned

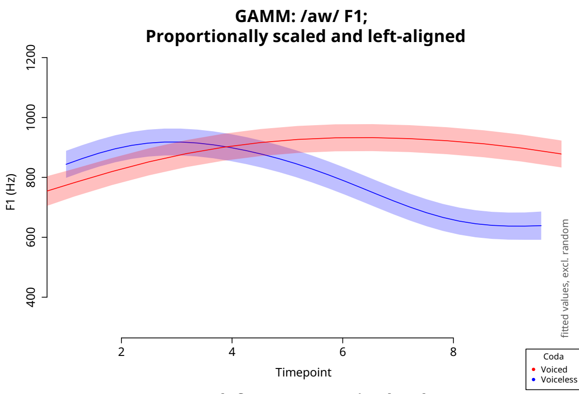


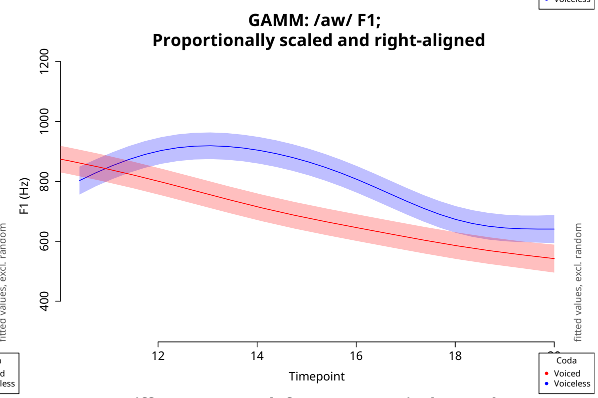
Difference smooth for /aj/ F2; voiceless coda Proportionally scaled and right-aligned

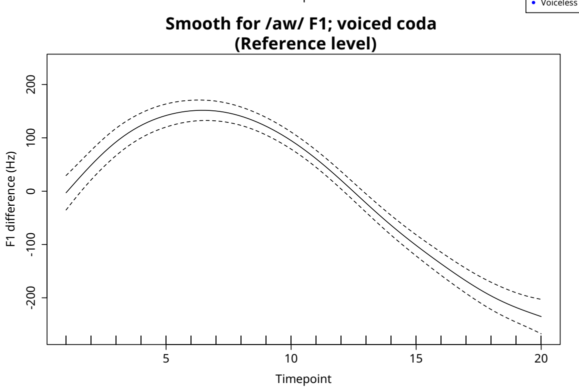


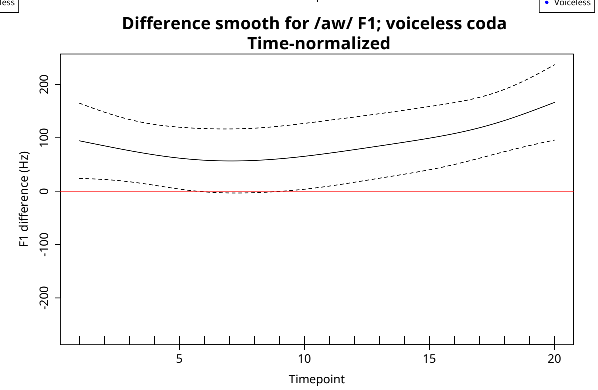
aw F1

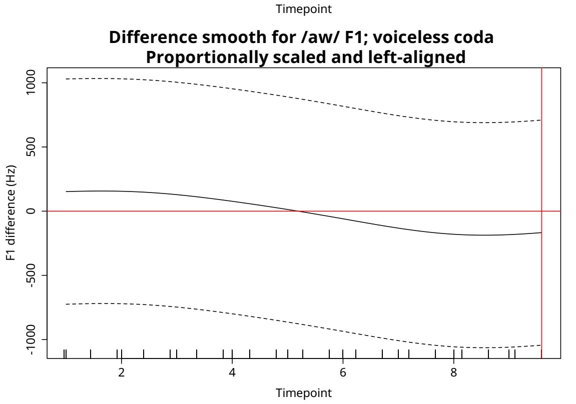

GAMM: /aw/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

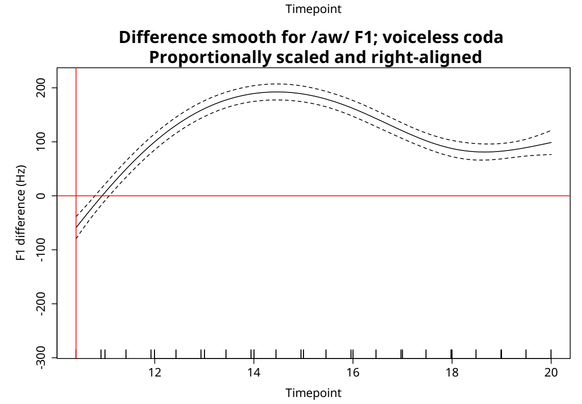

GAMM: /aw/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /aw/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

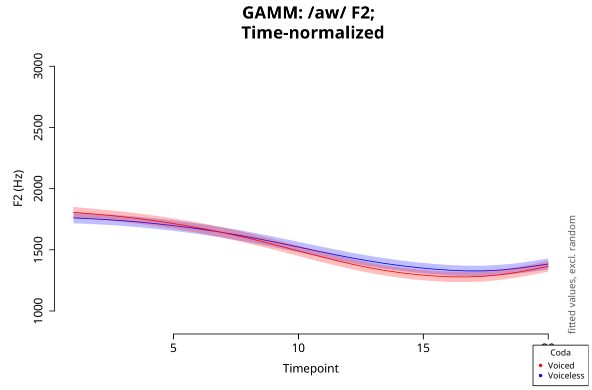

Smooth for /aw/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint

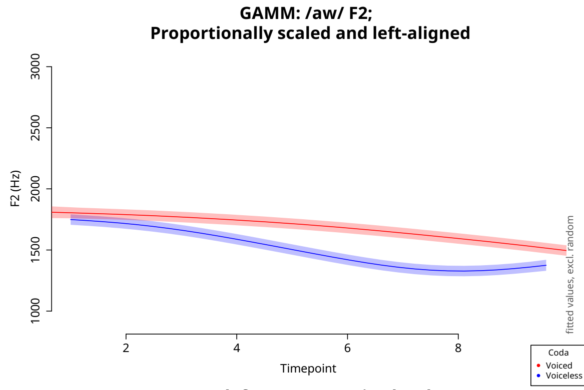

Difference smooth for /aw/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint

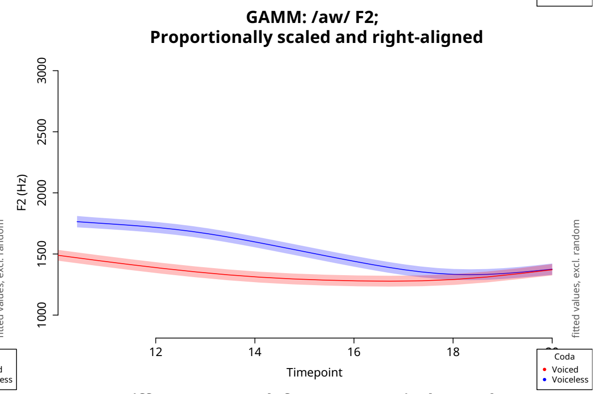

Difference smooth for /aw/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

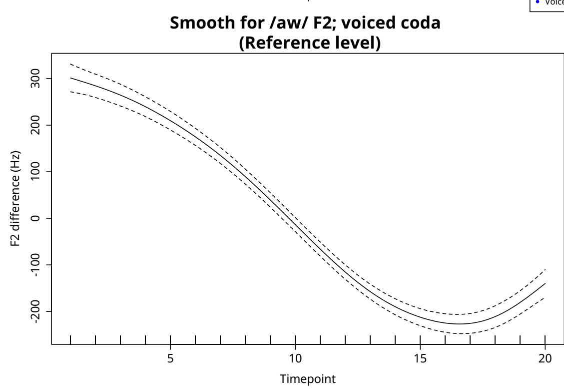

Difference smooth for /aw/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

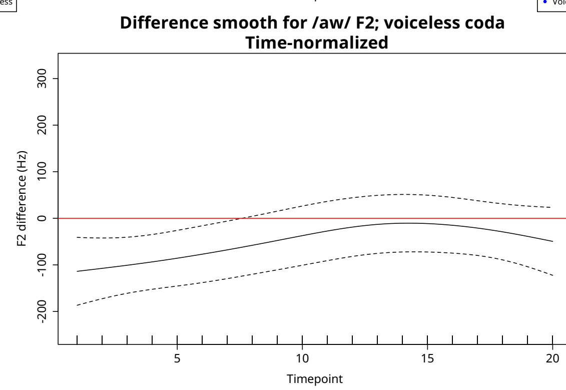
aw F2

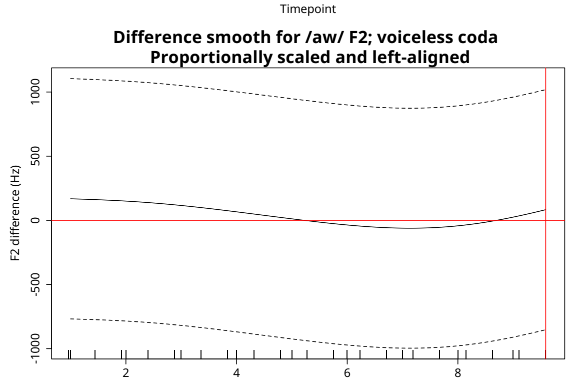

GAMM: /aw/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

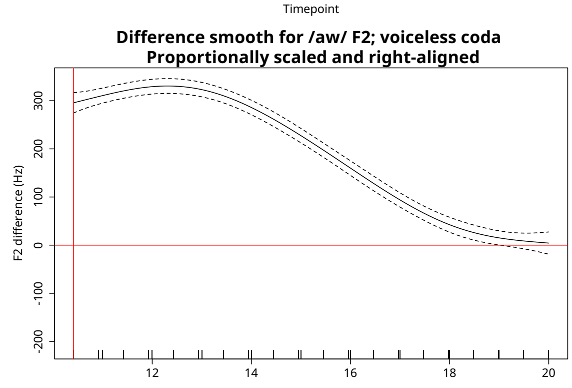

GAMM: /aw/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /aw/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

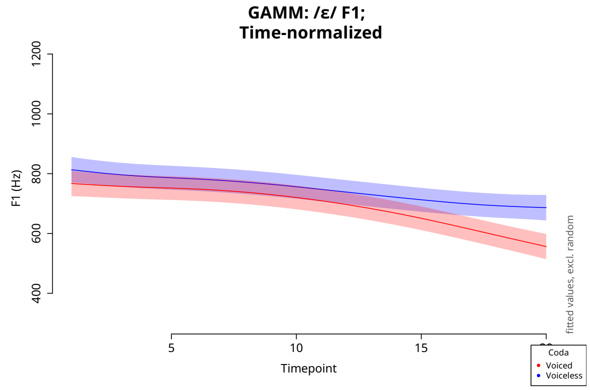

Smooth for /aw/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint

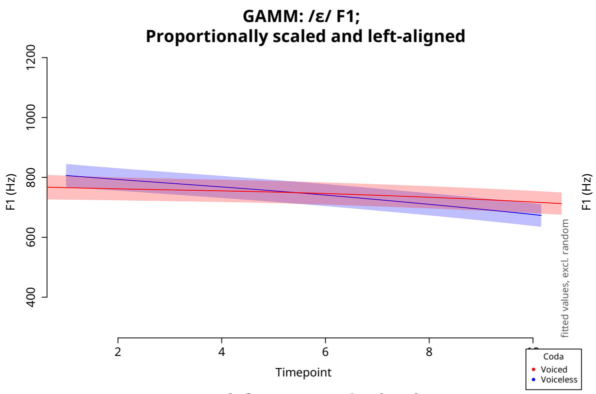

Difference smooth for /aw/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint

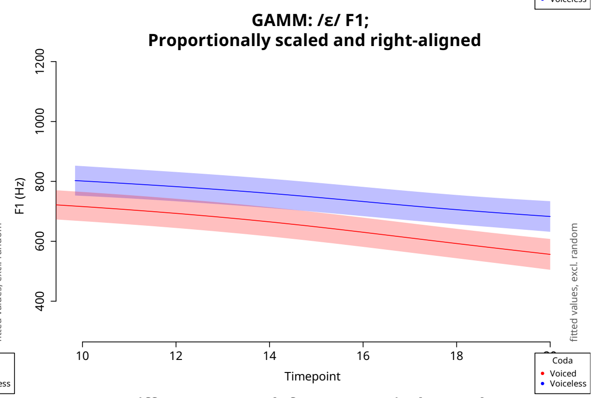

Difference smooth for /aw/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

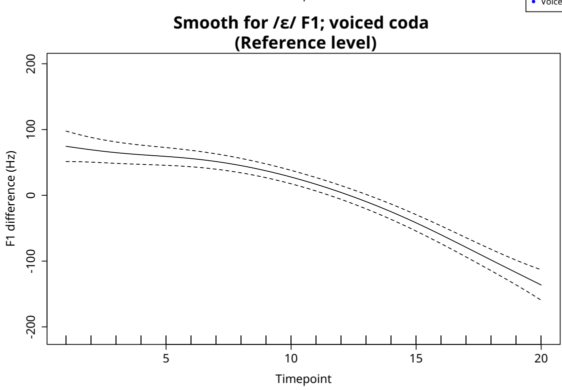

Difference smooth for /aw/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

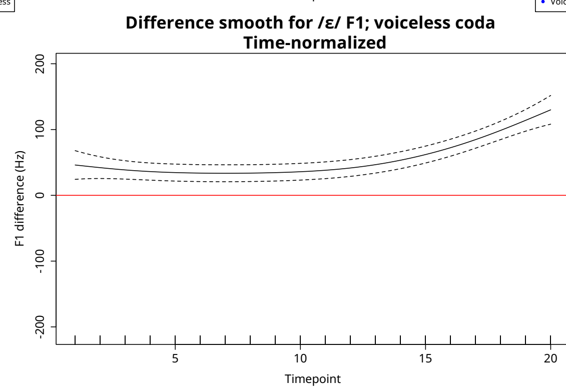
ε F1

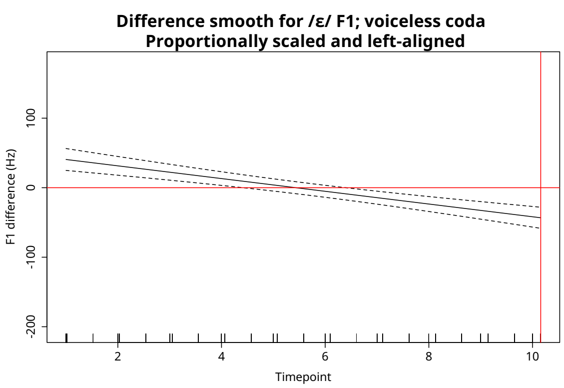

GAMM: /ɛ/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

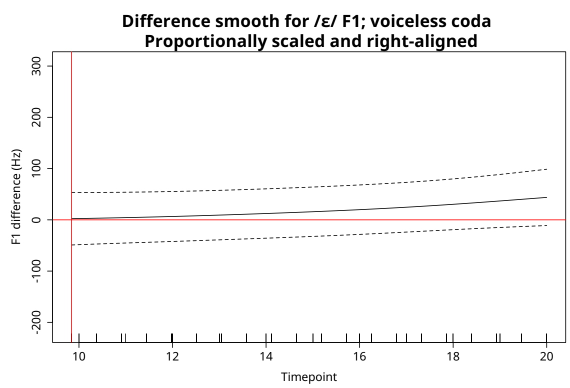

GAMM: /ɛ/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ɛ/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

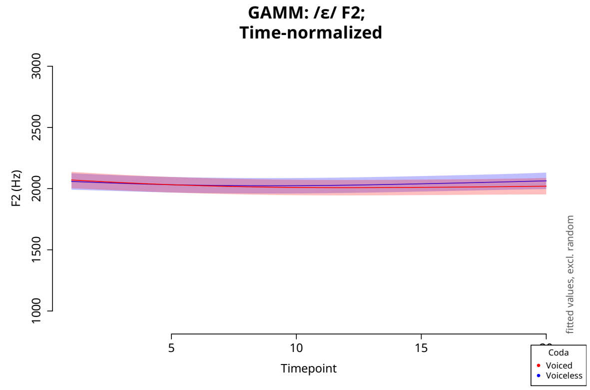

Smooth for /ɛ/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint

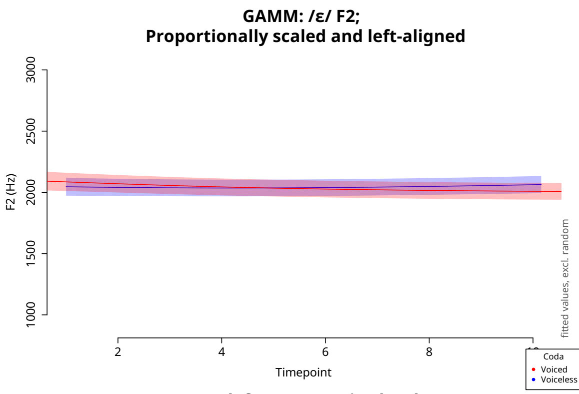

Difference smooth for /ɛ/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint

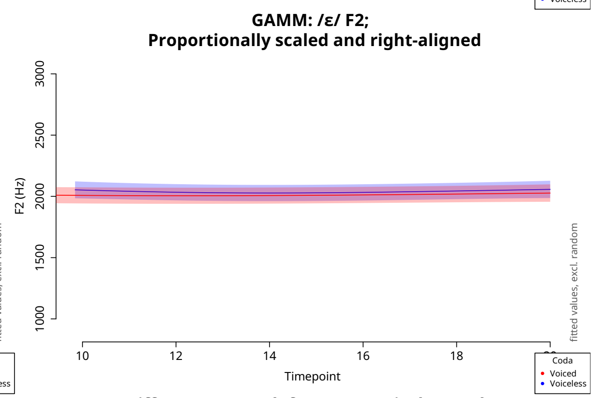

Difference smooth for /ɛ/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

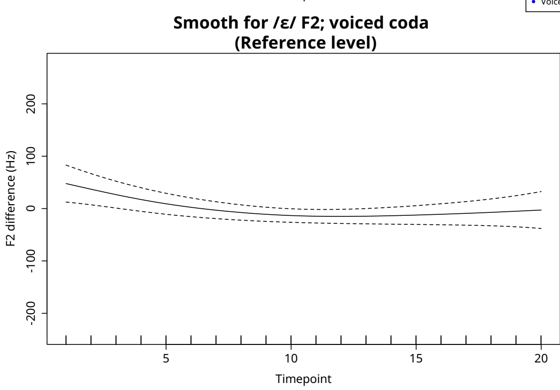

Difference smooth for /ɛ/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

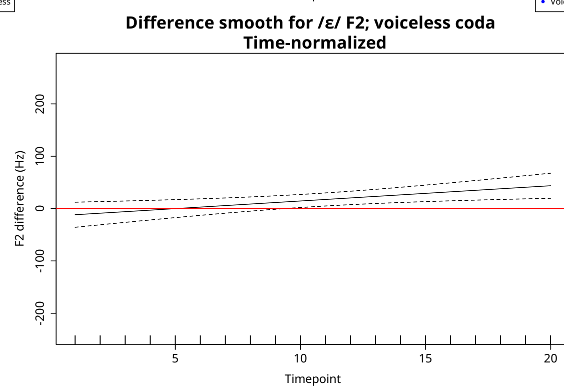
ɛ F2

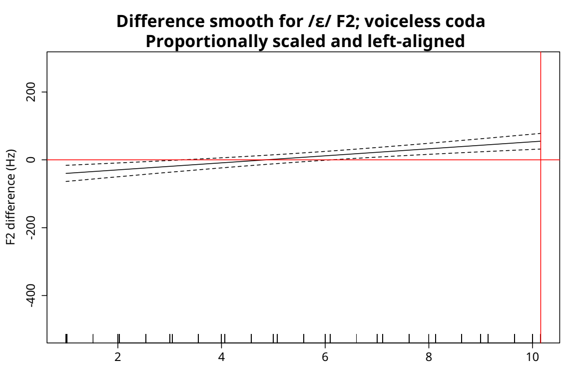

GAMM: /ɛ/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

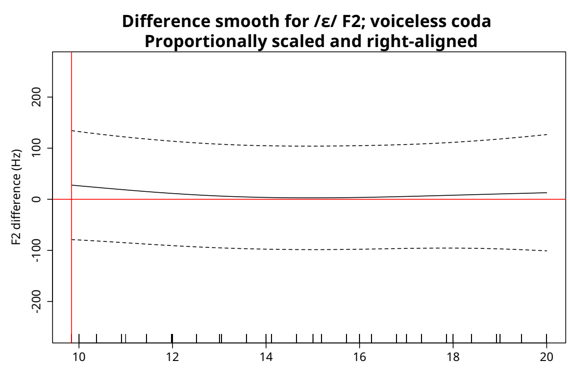

GAMM: /ɛ/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ɛ/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

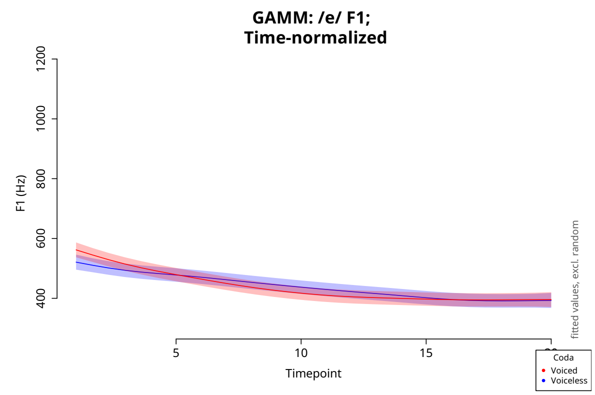

Smooth for /ɛ/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint

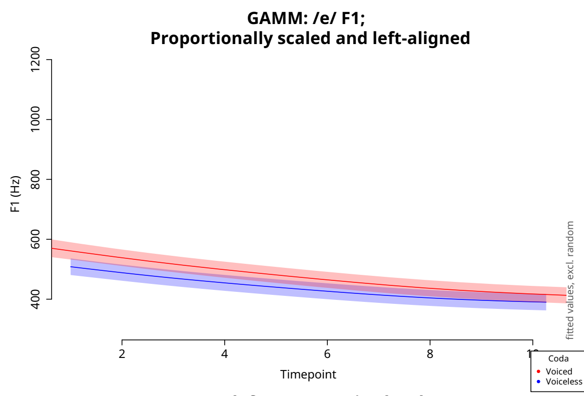

Difference smooth for /ɛ/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint

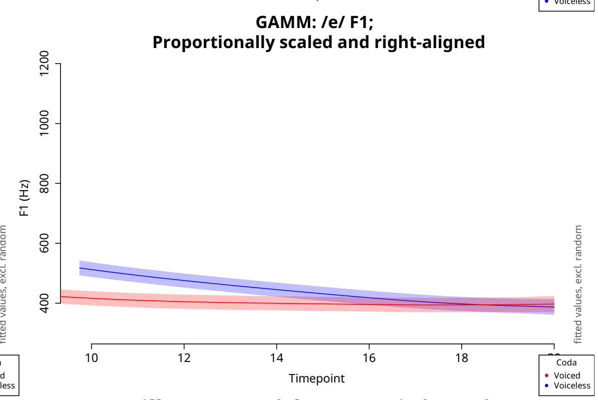

Difference smooth for /ɛ/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

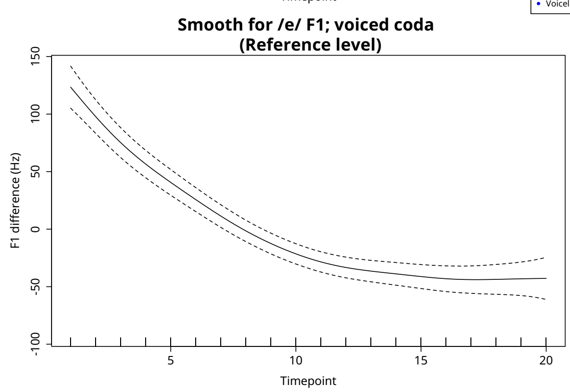

Difference smooth for /ɛ/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

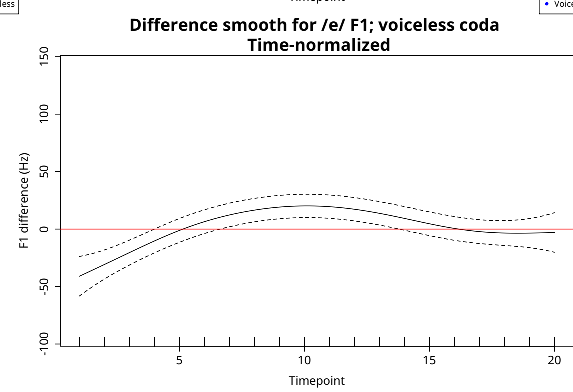
e F1

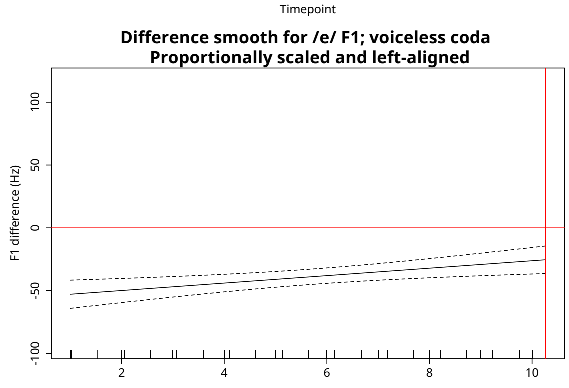

GAMM: /e/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

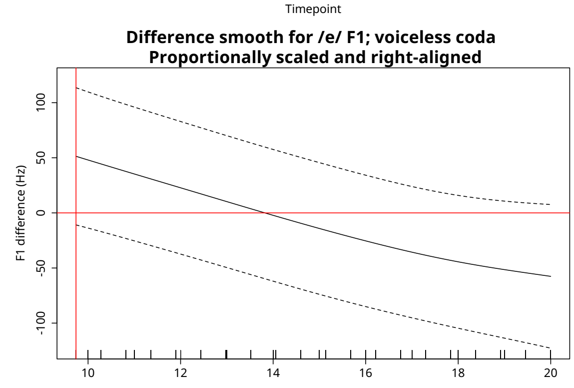

GAMM: /e/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /e/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

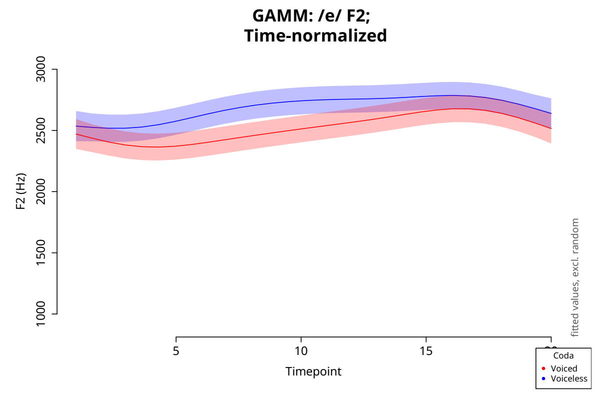

Smooth for /e/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint

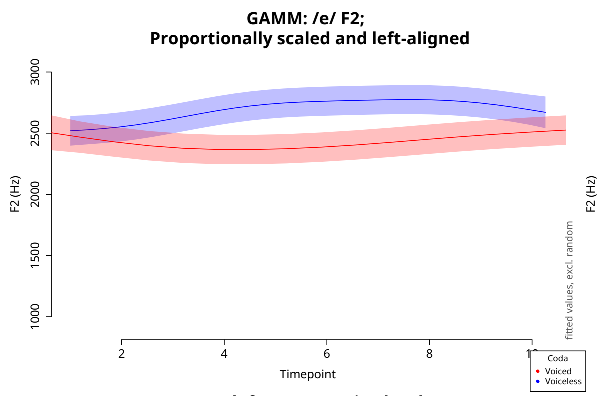

Difference smooth for /e/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint

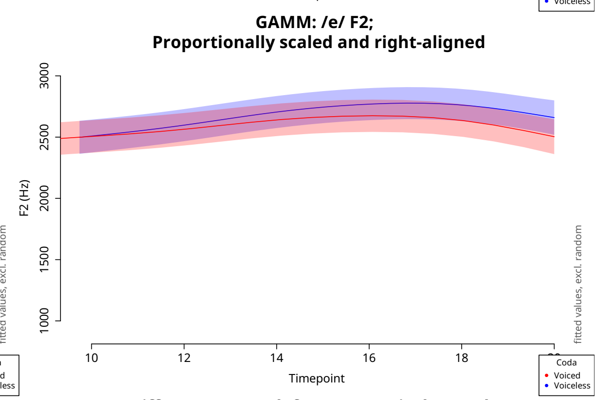

Difference smooth for /e/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

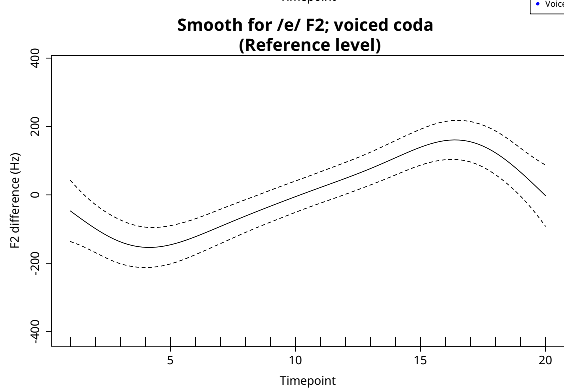

Difference smooth for /e/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

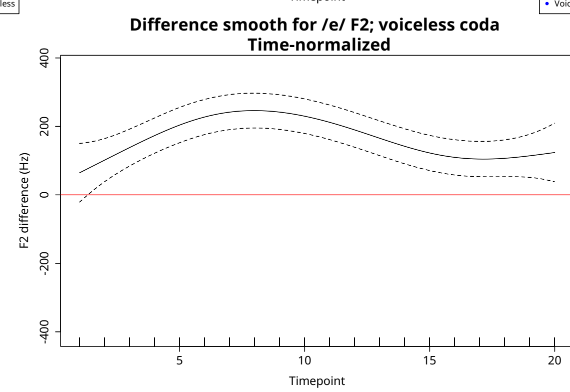
e F2

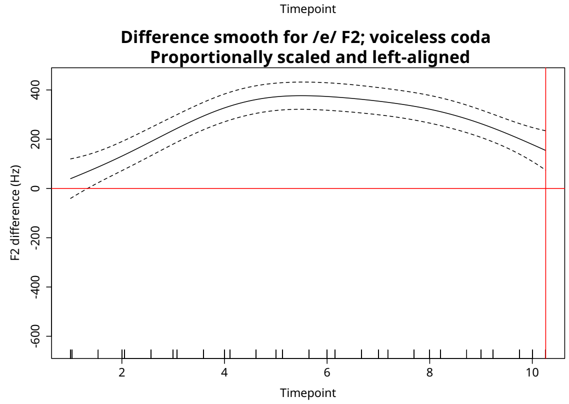

GAMM: /e/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

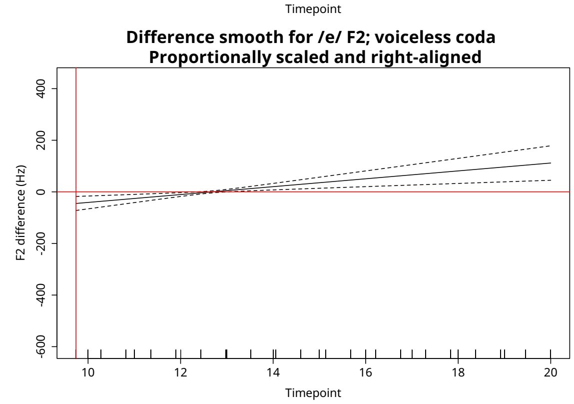

GAMM: /e/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /e/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

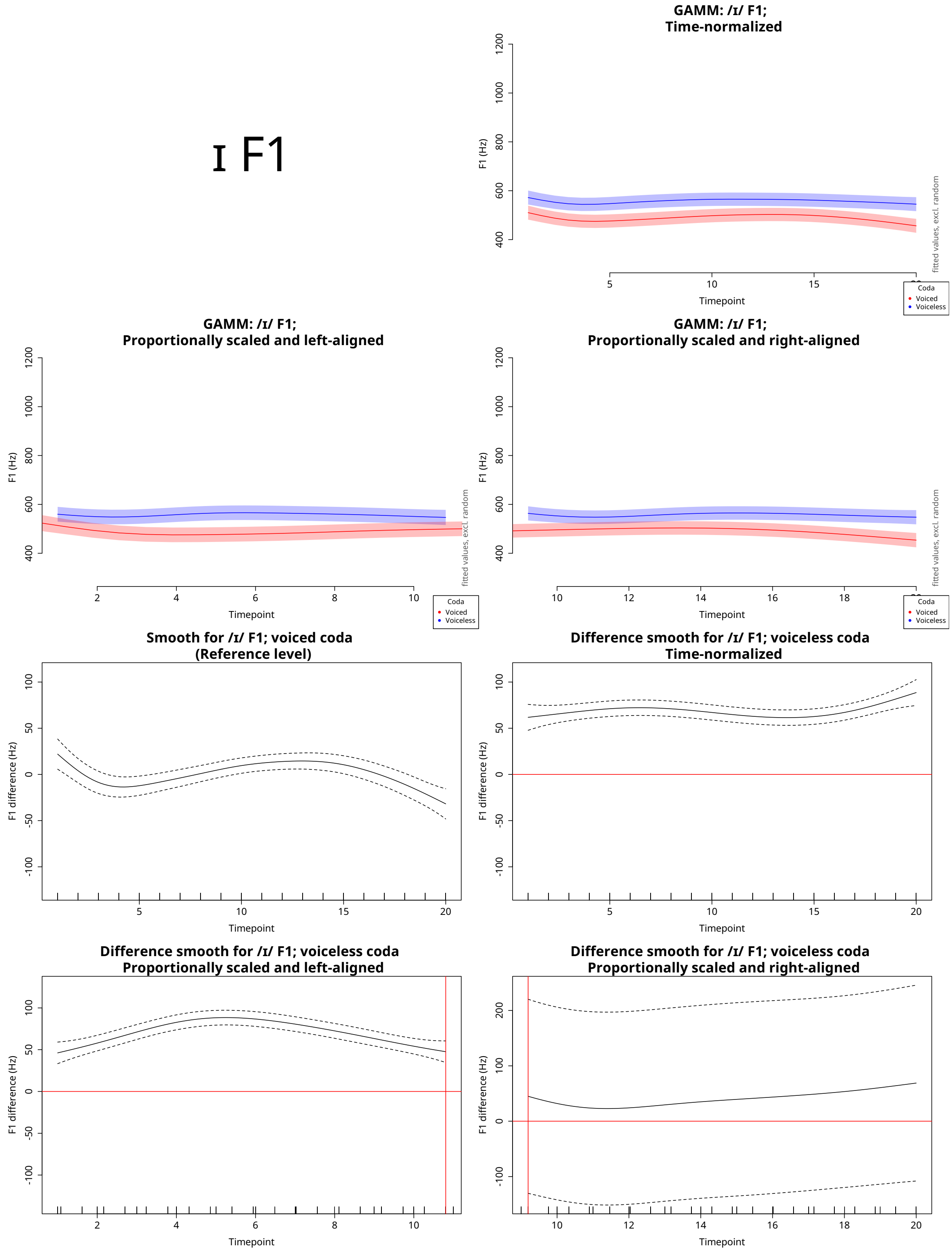

Smooth for /e/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint


Difference smooth for /e/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint

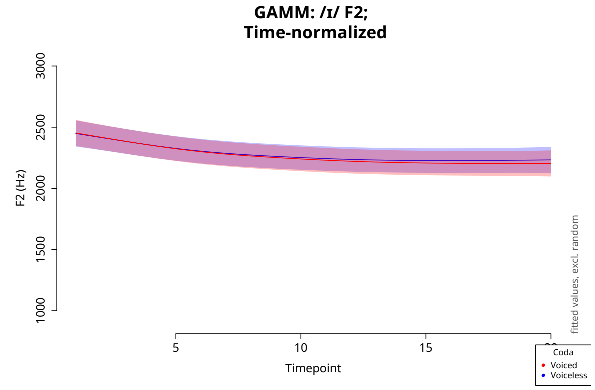

Difference smooth for /e/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

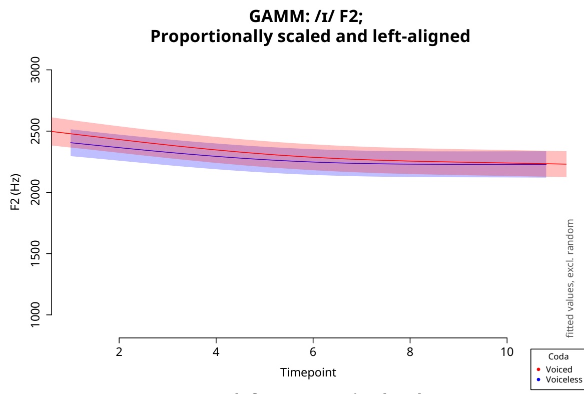

Difference smooth for /e/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

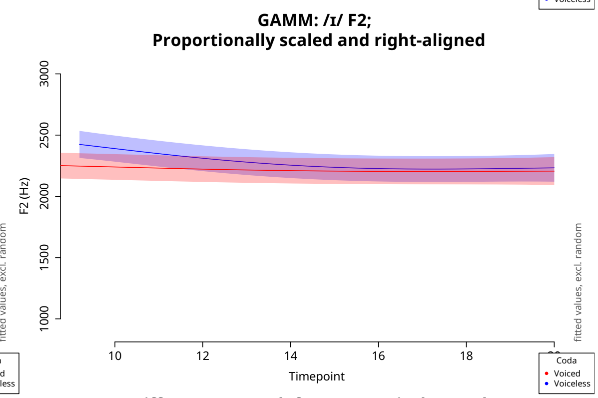
ɪ F1

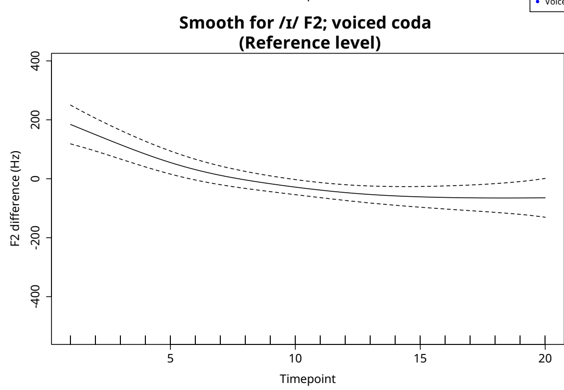

GAMM: /ɪ/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless
GAMM: /ɪ/ F1;
Proportionally scaled and left-aligned
GAMM: /ɪ/ F1;
Proportionally scaled and right-aligned
Smooth for /ɪ/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Difference smooth for /ɪ/ F1; voiceless coda
Time-normalized
Difference smooth for /ɪ/ F1; voiceless coda
Proportionally scaled and left-aligned
Difference smooth for /ɪ/ F1; voiceless coda
Proportionally scaled and right-aligned

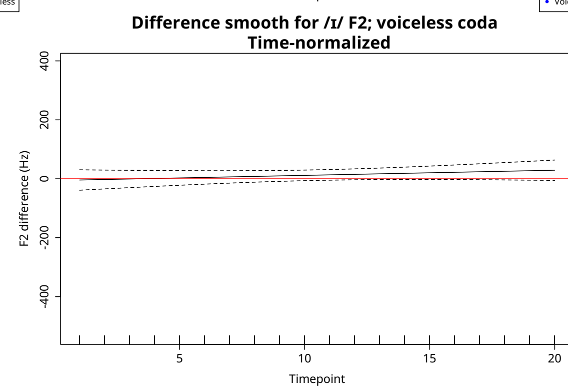
ɪ F2

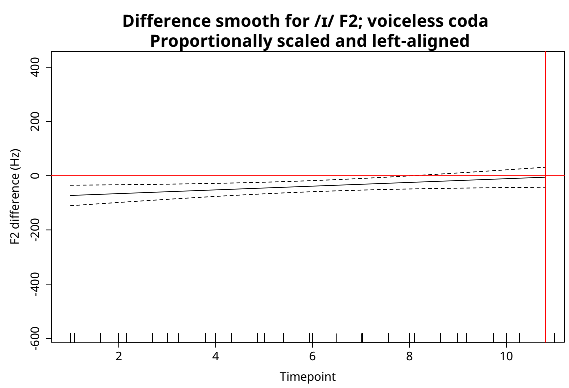

GAMM: /ɪ/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

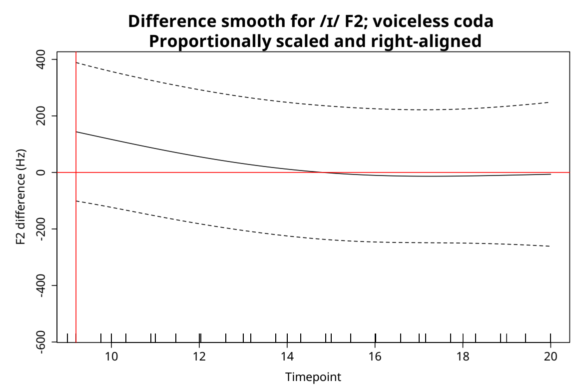

GAMM: /ɪ/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ɪ/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


Smooth for /ɪ/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint

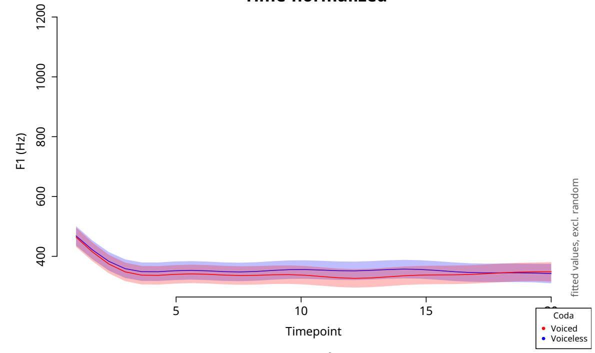

Difference smooth for /ɪ/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint


Difference smooth for /ɪ/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

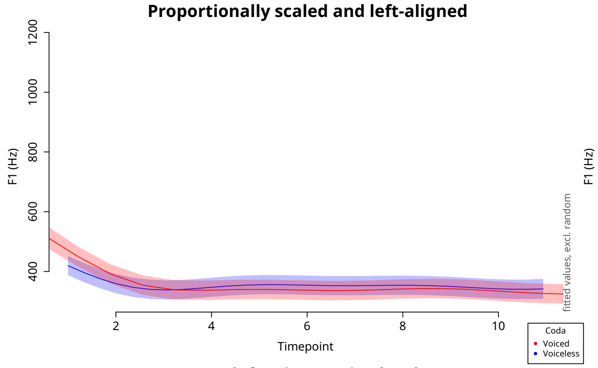

Difference smooth for /ɪ/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

i F1

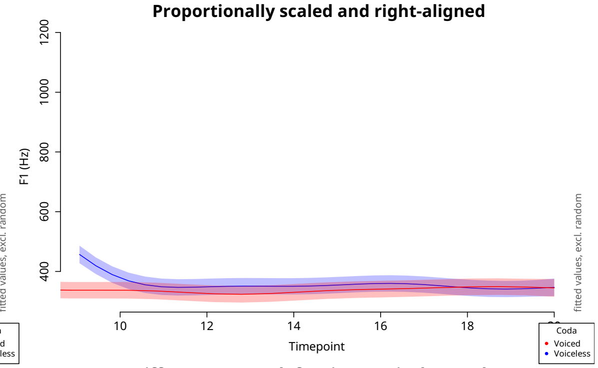
GAMM: /i/ F1; Time-normalized



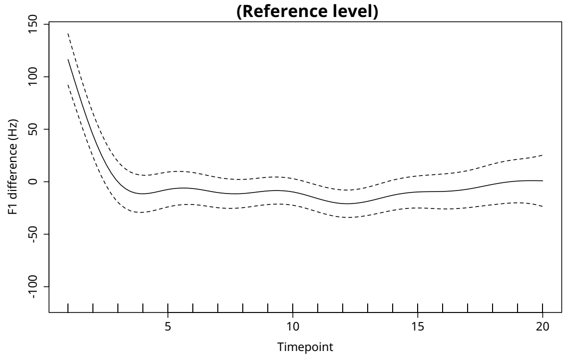
GAMM: /i/ F1; Proportionally scaled and left-aligned



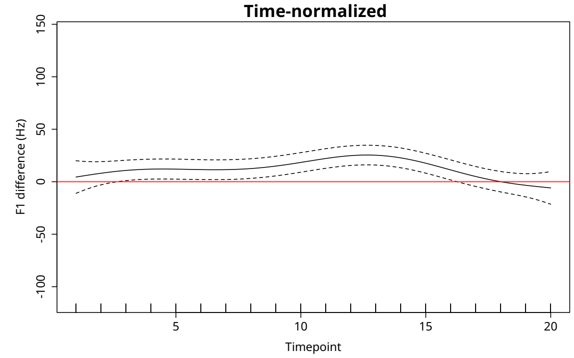
GAMM: /i/ F1; Proportionally scaled and right-aligned



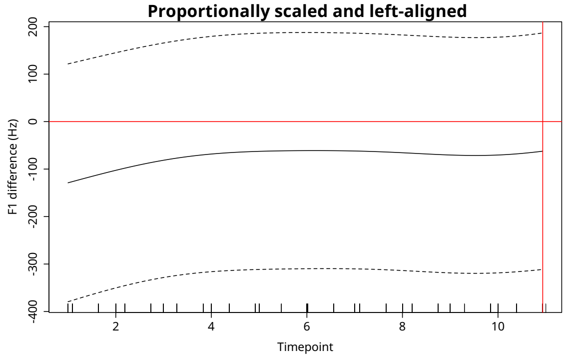
Smooth for /i/ F1; voiced coda (Reference level)



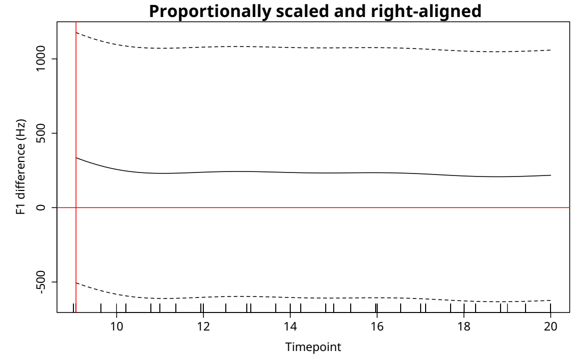
Difference smooth for /i/ F1; voiceless coda Time-normalized



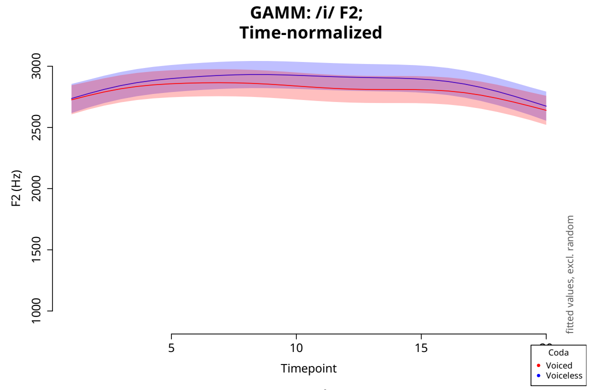
Difference smooth for /i/ F1; voiceless coda Proportionally scaled and left-aligned

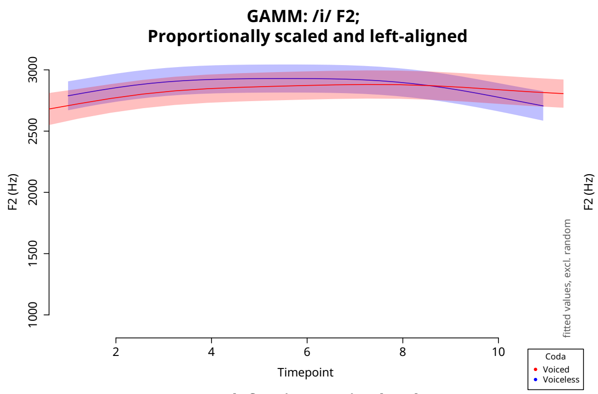


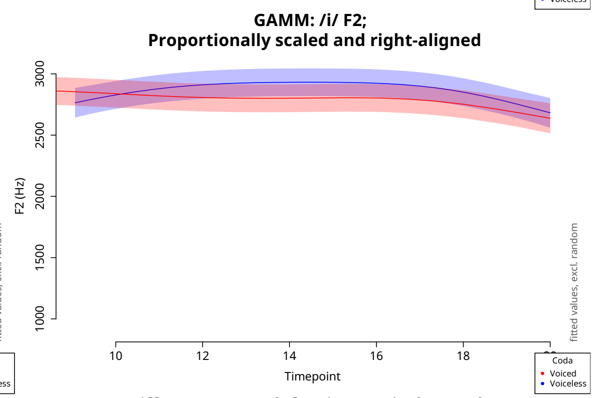
Difference smooth for /i/ F1; voiceless coda Proportionally scaled and right-aligned

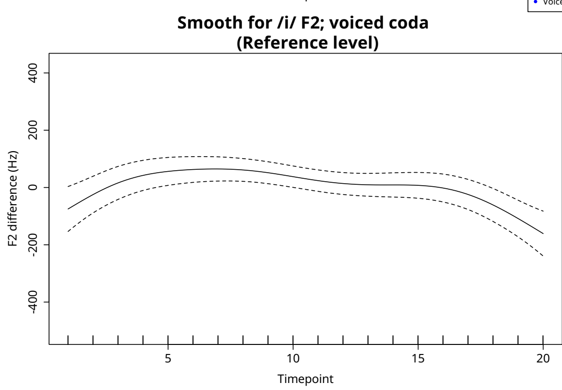


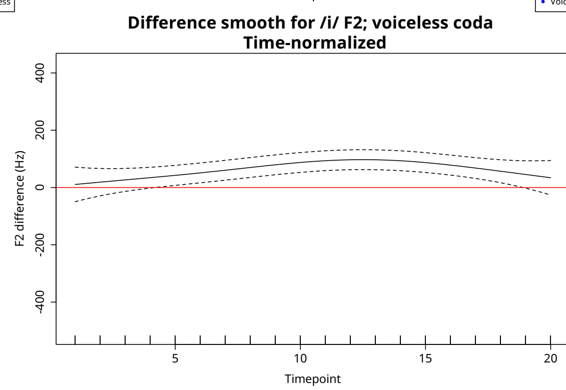
i F2

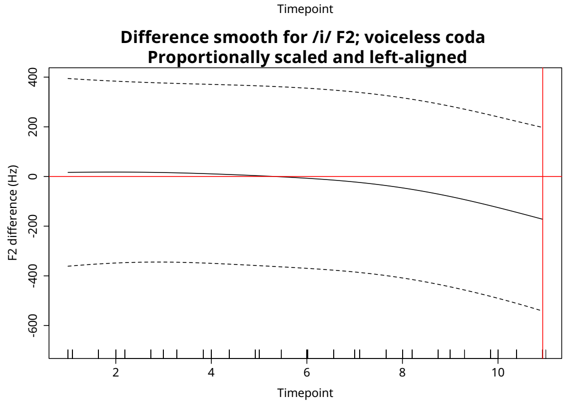

GAMM: /i/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

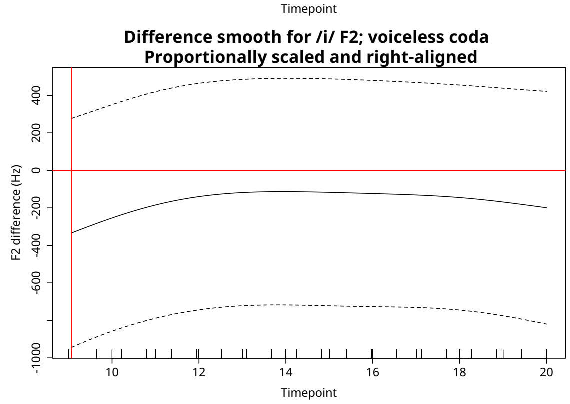

GAMM: /i/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /i/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

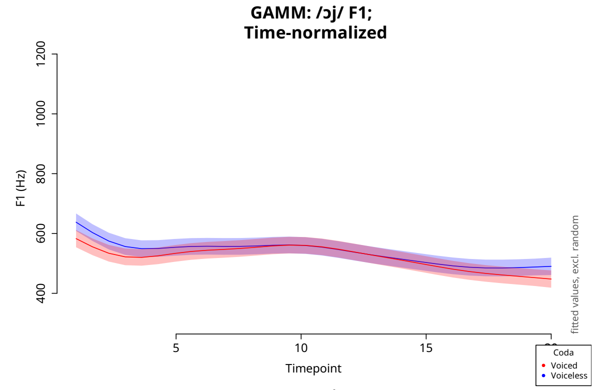

Smooth for /i/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint

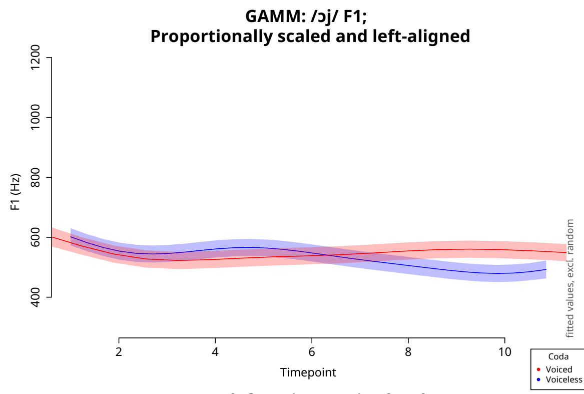

Difference smooth for /i/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint

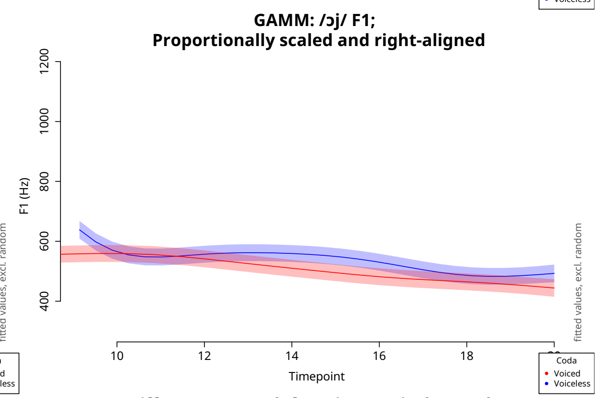

Difference smooth for /i/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

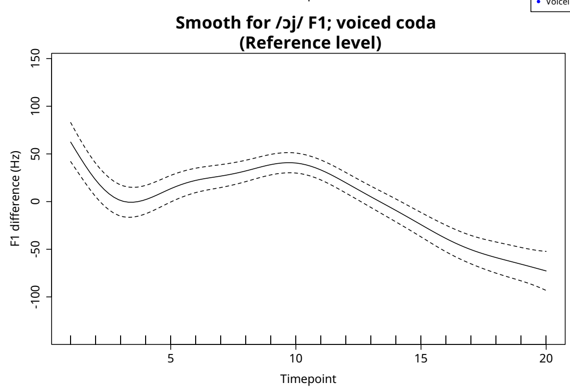

Difference smooth for /i/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

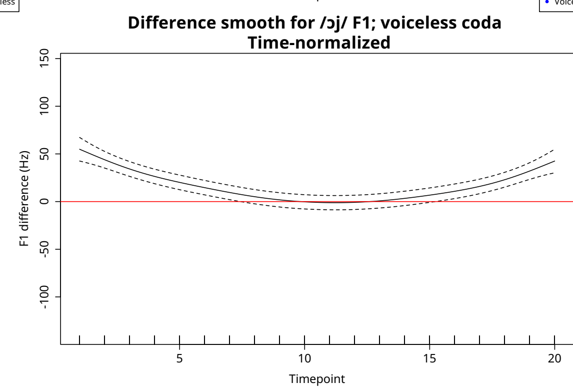
ɔj F1

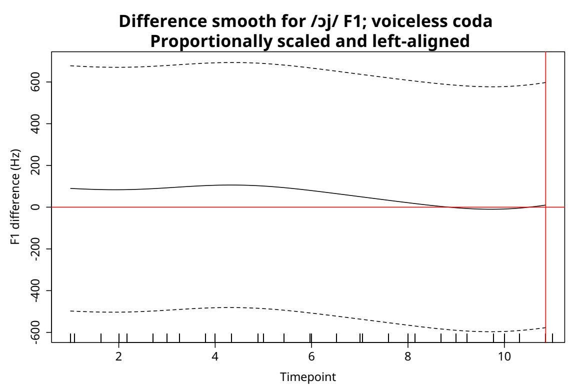

GAMM: /ɔj/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

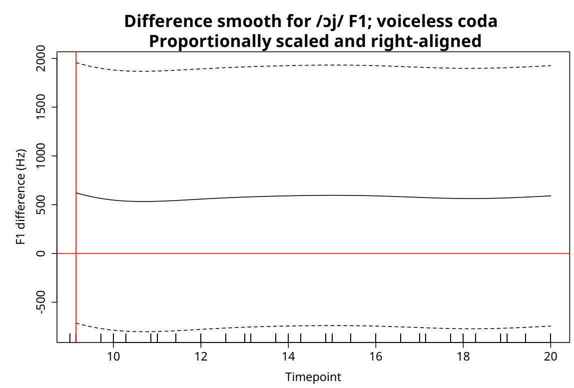

GAMM: /ɔj/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ɔj/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

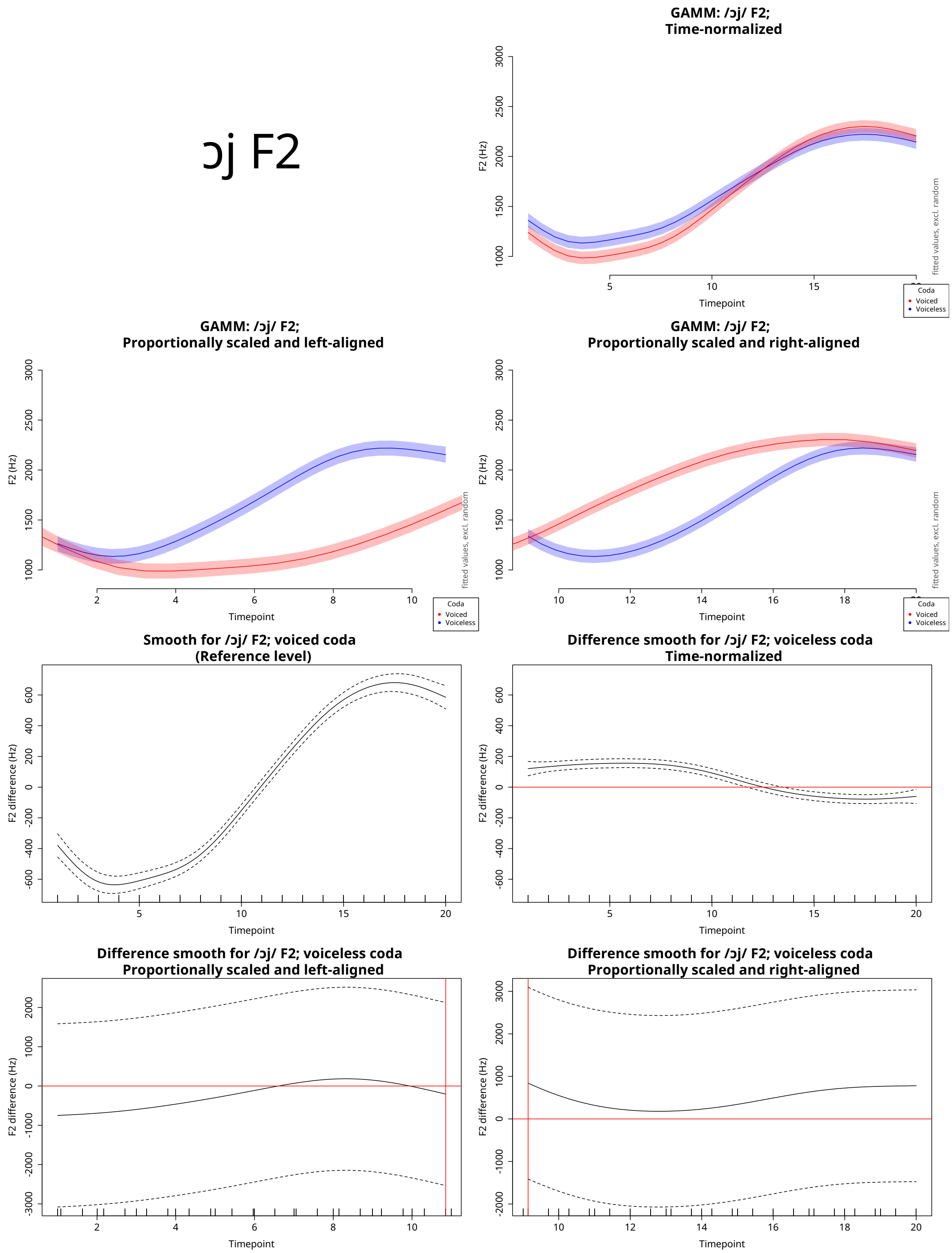

Smooth for /ɔj/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint


Difference smooth for /ɔj/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint

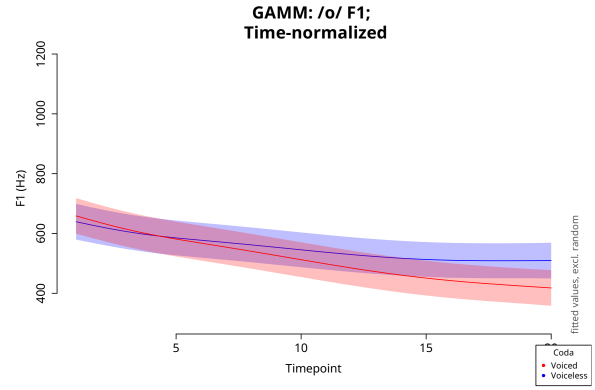

Difference smooth for /ɔj/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

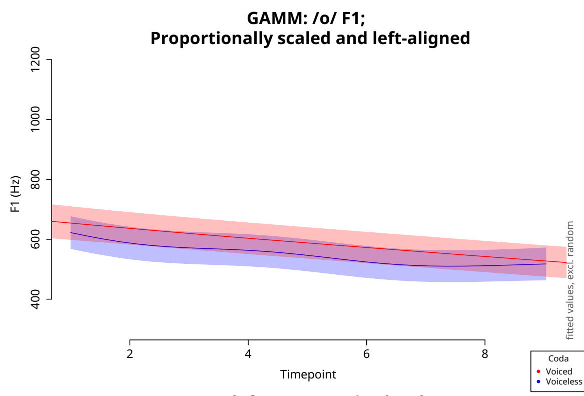

Difference smooth for /ɔj/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

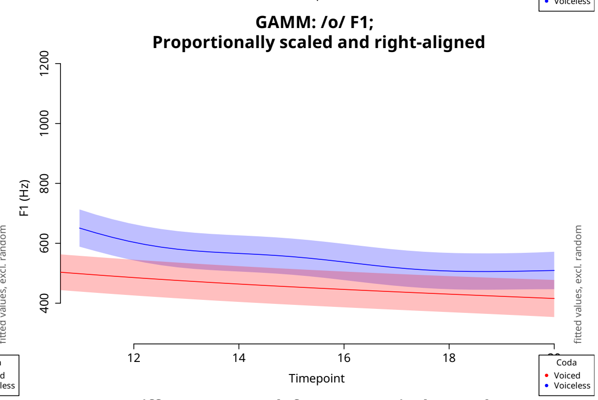
ɔj F2

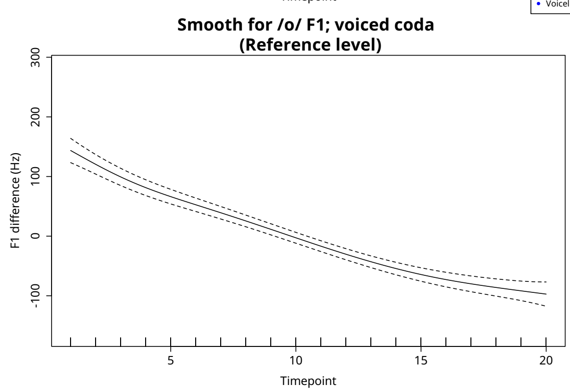

GAMM: /ɔj/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless
GAMM: /ɔj/ F2;
Proportionally scaled and left-aligned
GAMM: /ɔj/ F2;
Proportionally scaled and right-aligned
Smooth for /ɔj/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Difference smooth for /ɔj/ F2; voiceless coda
Time-normalized
Difference smooth for /ɔj/ F2; voiceless coda
Proportionally scaled and left-aligned
Difference smooth for /ɔj/ F2; voiceless coda
Proportionally scaled and right-aligned

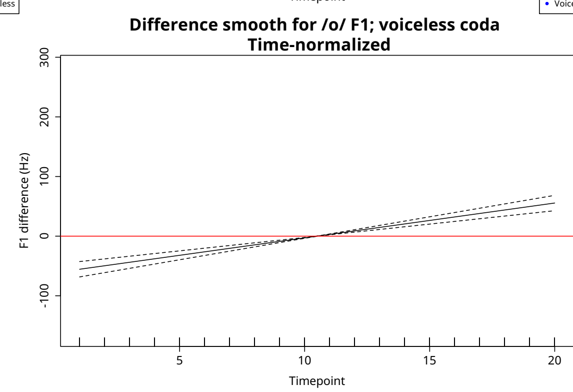
o F1

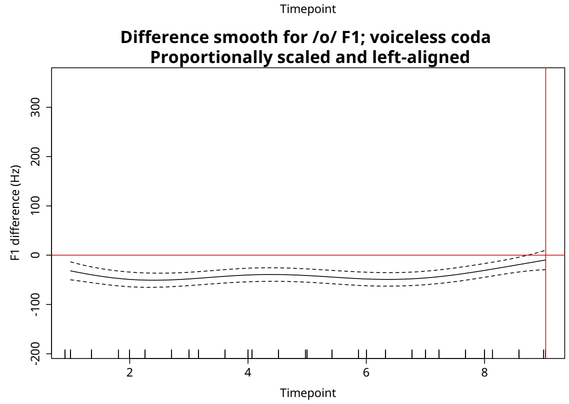

GAMM: /o/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

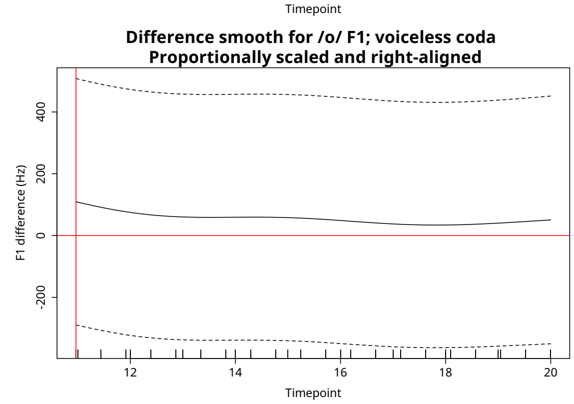

GAMM: /o/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /o/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

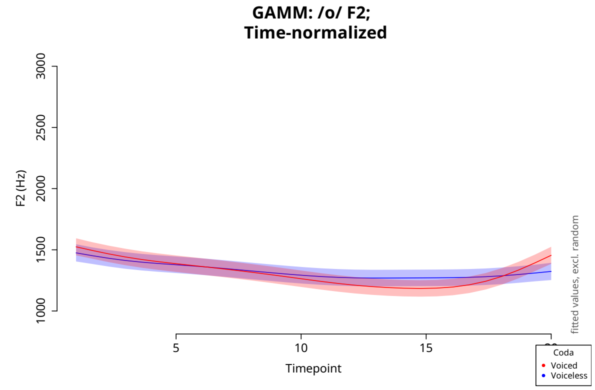

Smooth for /o/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint

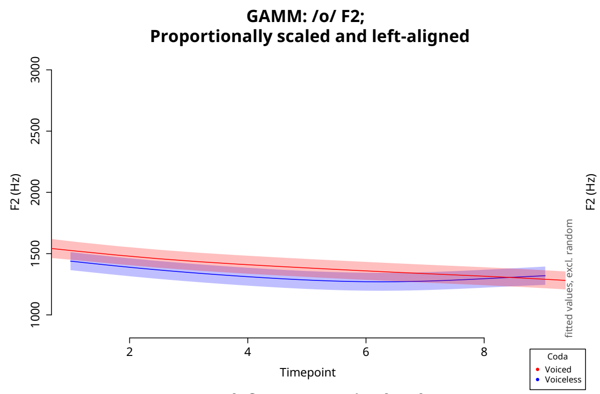

Difference smooth for /o/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint

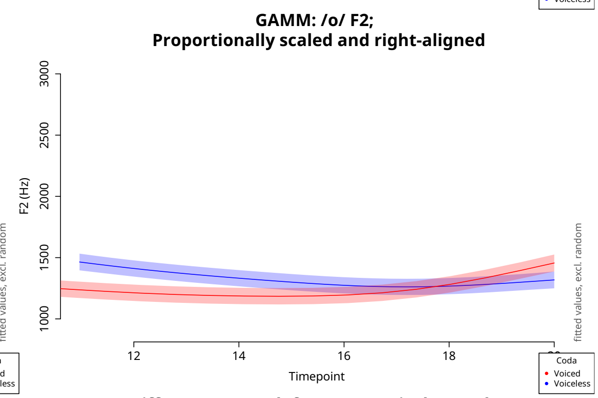

Difference smooth for /o/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

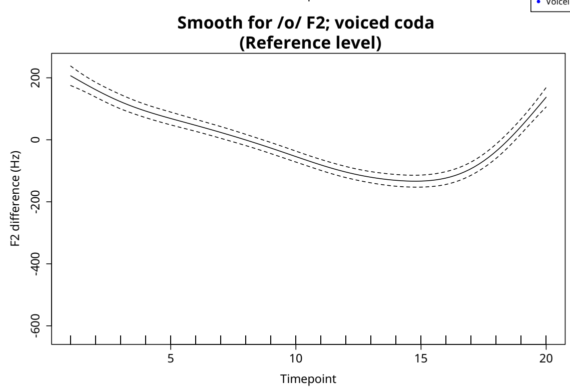

Difference smooth for /o/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

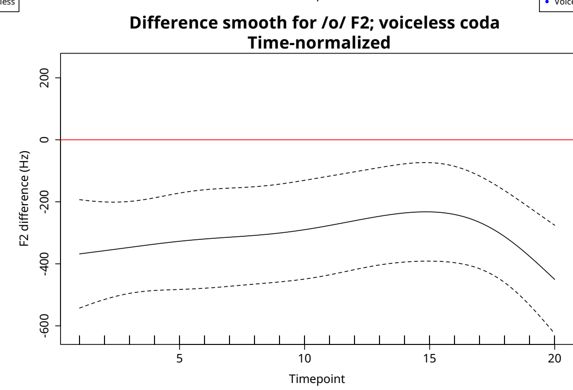
o F2

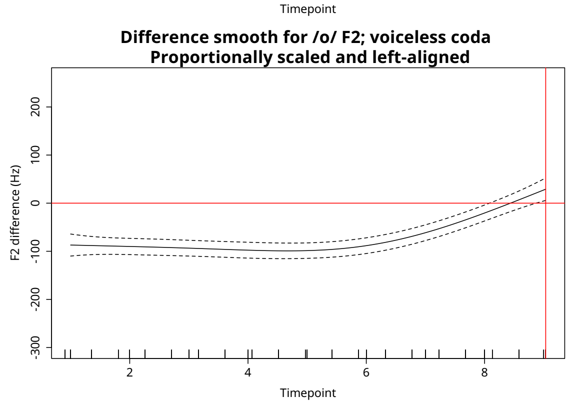

GAMM: /o/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

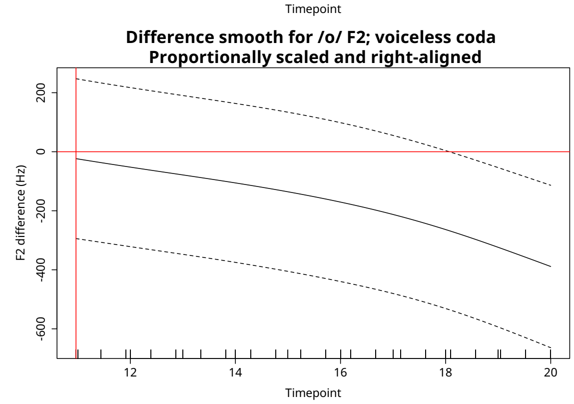

GAMM: /o/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /o/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

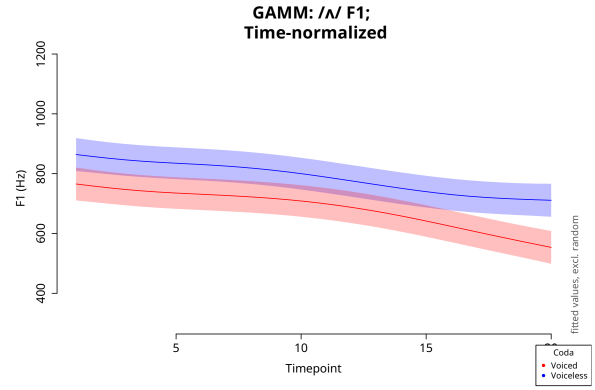

Smooth for /o/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint

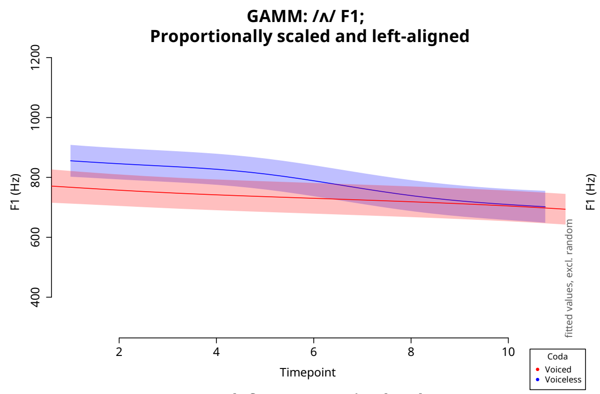

Difference smooth for /o/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint

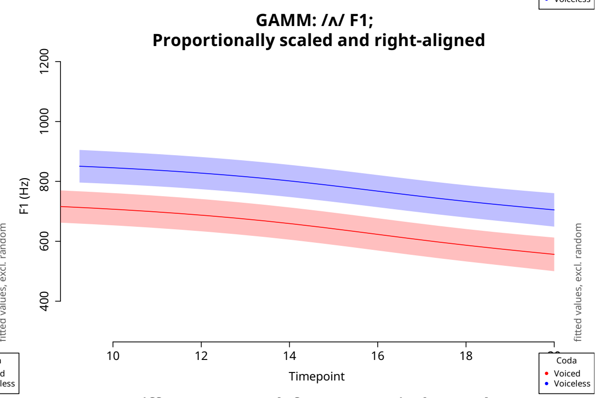

Difference smooth for /o/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

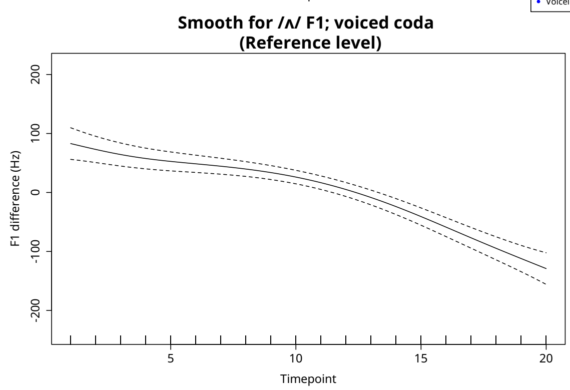

Difference smooth for /o/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

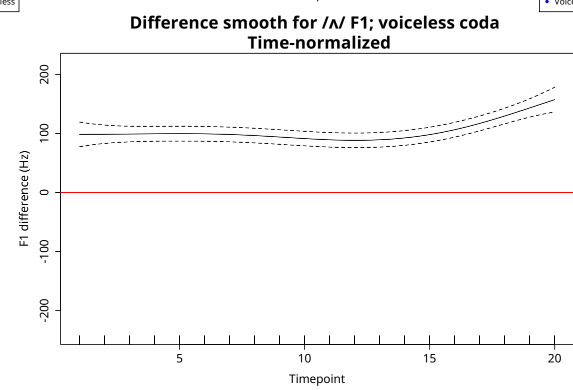
ʌ F1

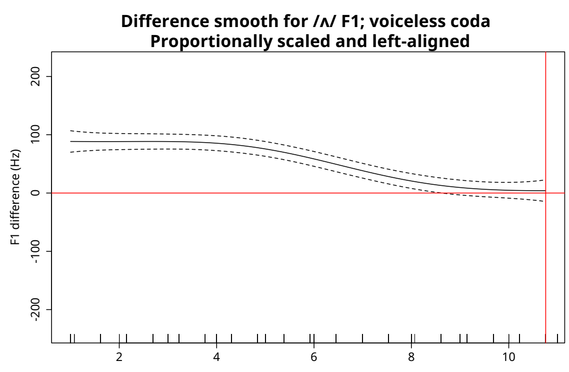

GAMM: /ʌ/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

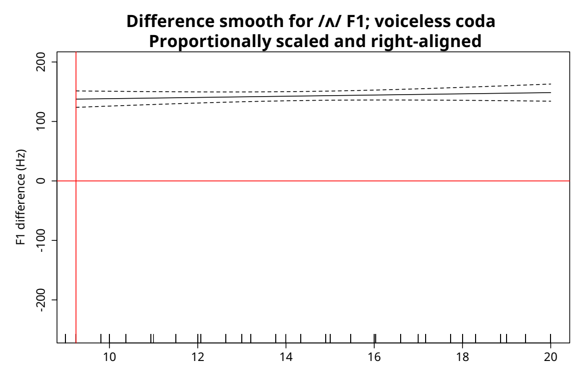

GAMM: /ʌ/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ʌ/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

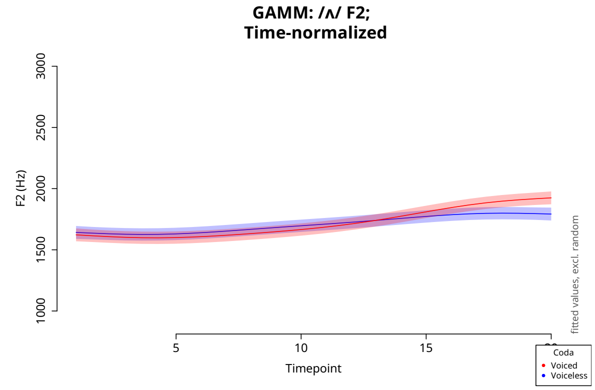

Smooth for /ʌ/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint

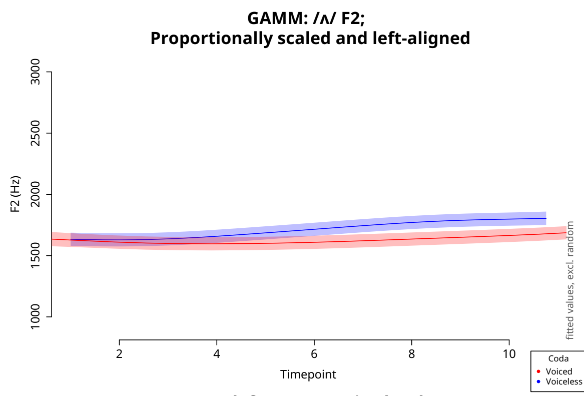

Difference smooth for /ʌ/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint

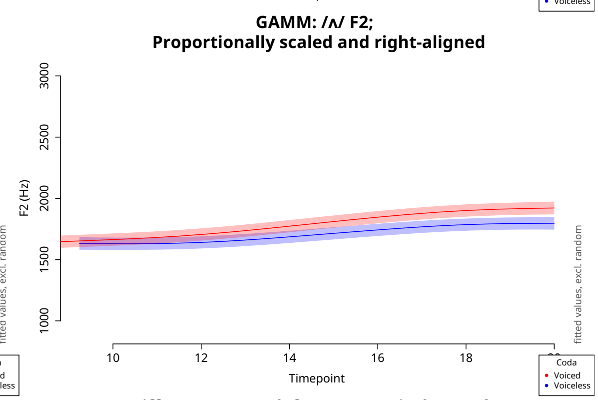

Difference smooth for /ʌ/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

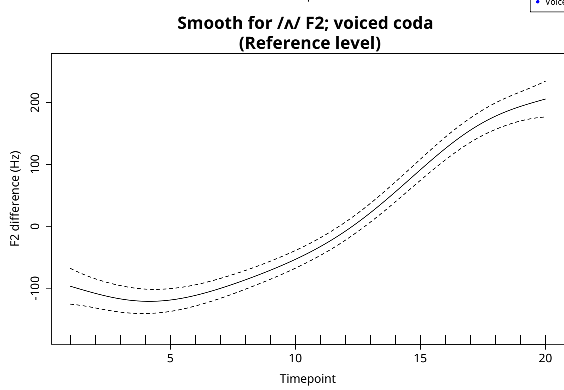

Difference smooth for /ʌ/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

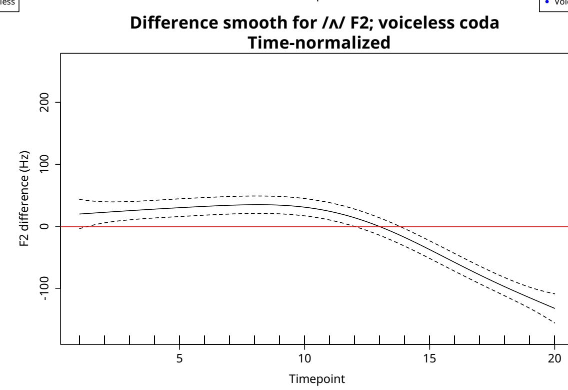
ʌ F2

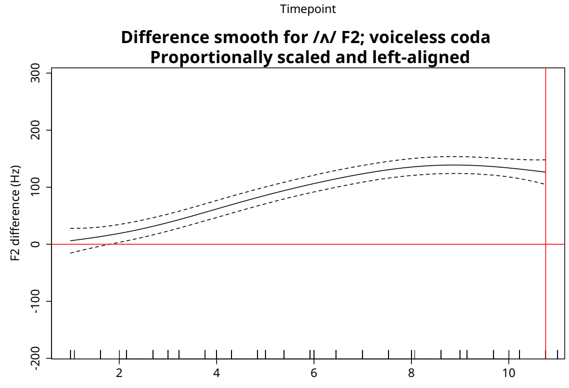

GAMM: /ʌ/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

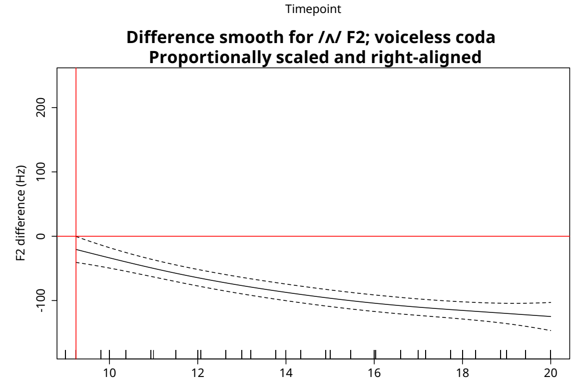

GAMM: /ʌ/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ʌ/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


Smooth for /ʌ/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint

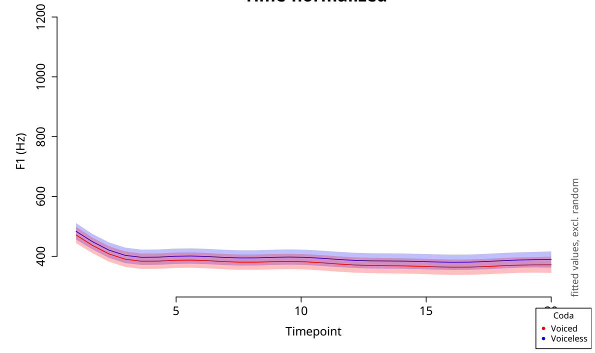

Difference smooth for /ʌ/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint


Difference smooth for /ʌ/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

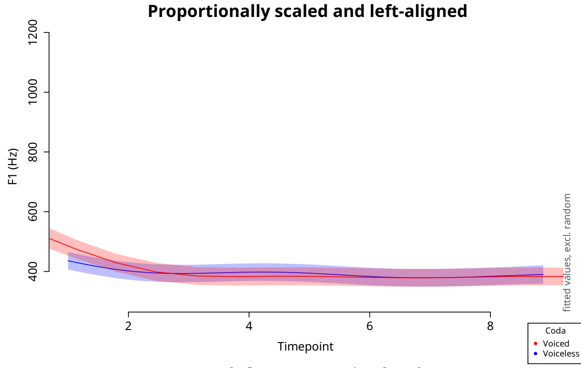

Difference smooth for /ʌ/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

u F1

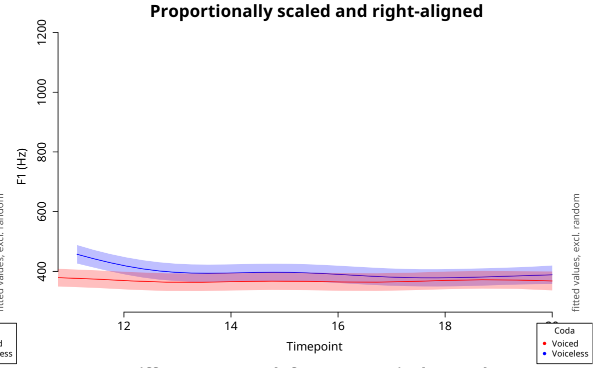
**GAMM: /u/ F1;
Time-normalized**



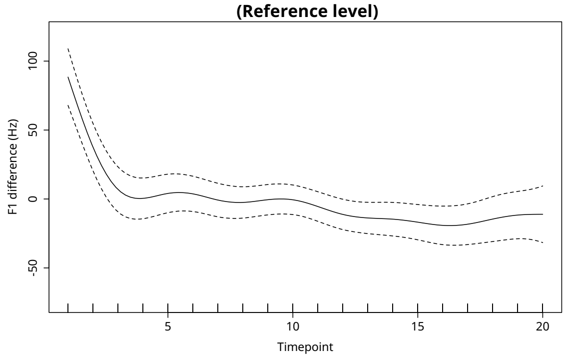
**GAMM: /u/ F1;
Proportionally scaled and left-aligned**



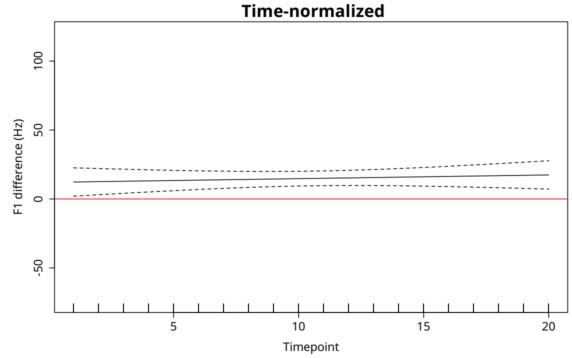
**GAMM: /u/ F1;
Proportionally scaled and right-aligned**



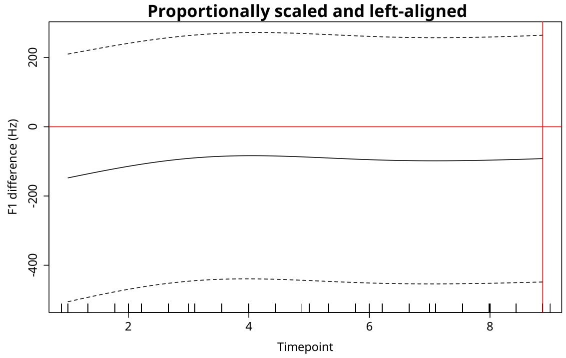
**Smooth for /u/ F1; voiced coda
(Reference level)**



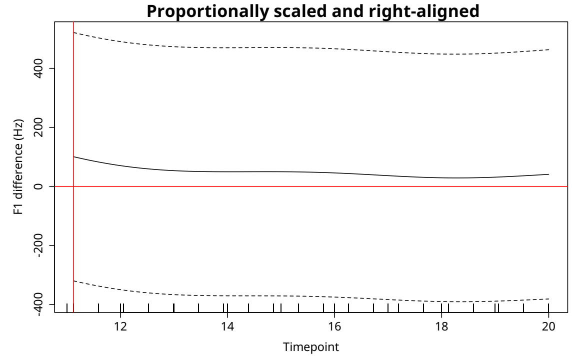
**Difference smooth for /u/ F1; voiceless coda
Time-normalized**



**Difference smooth for /u/ F1; voiceless coda
Proportionally scaled and left-aligned**

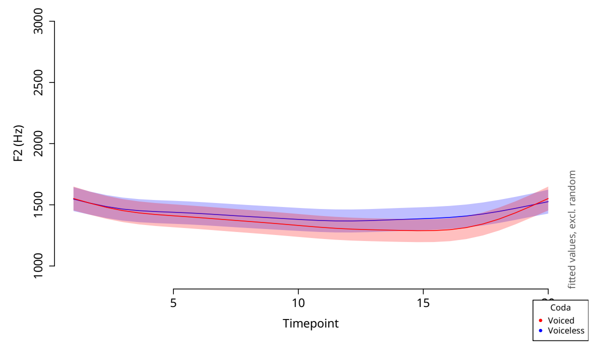


**Difference smooth for /u/ F1; voiceless coda
Proportionally scaled and right-aligned**

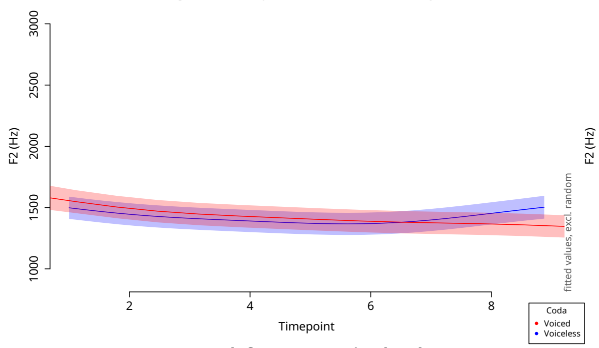


u F2

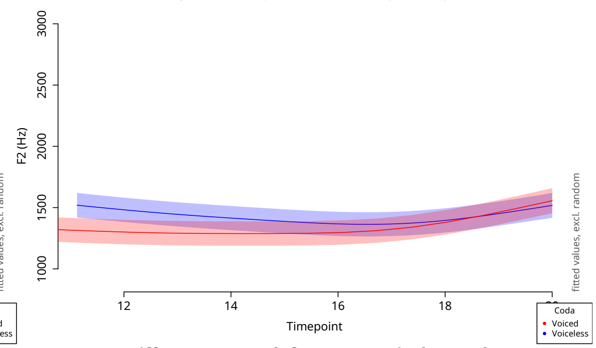
GAMM: /u/ F2;
Time-normalized



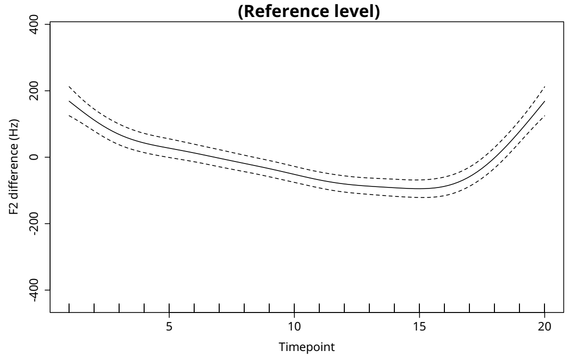
GAMM: /u/ F2;
Proportionally scaled and left-aligned



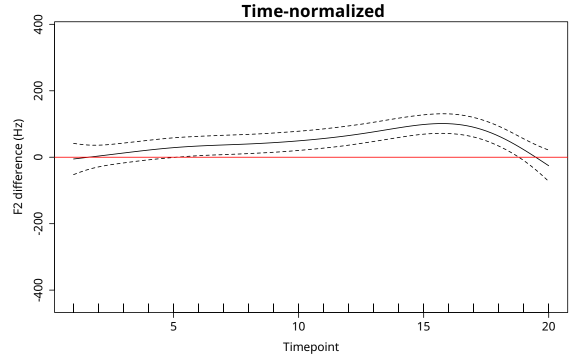
GAMM: /u/ F2;
Proportionally scaled and right-aligned



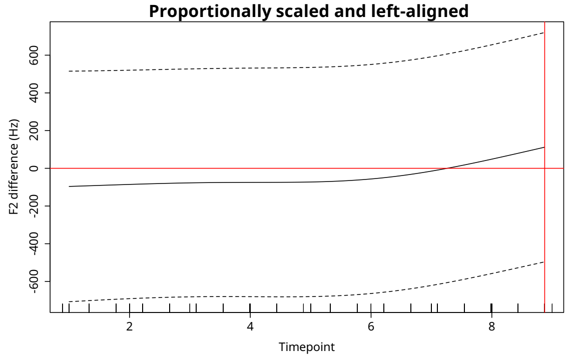
Smooth for /u/ F2; voiced coda
(Reference level)



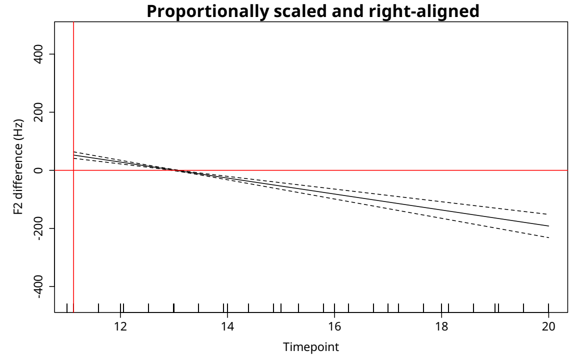
Difference smooth for /u/ F2; voiceless coda
Time-normalized



Difference smooth for /u/ F2; voiceless coda
Proportionally scaled and left-aligned



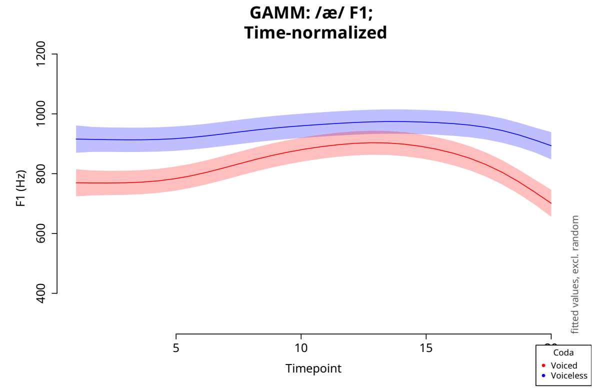
Difference smooth for /u/ F2; voiceless coda
Proportionally scaled and right-aligned



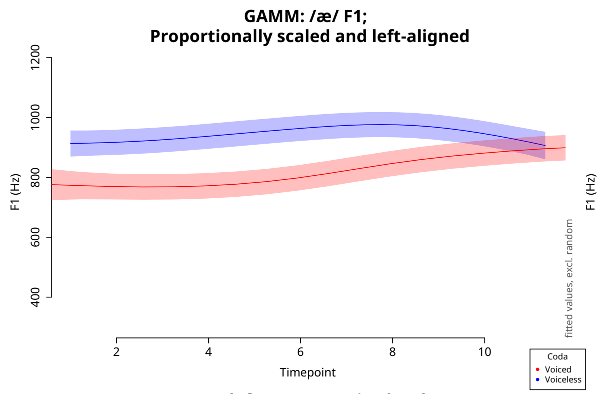
Appendix G GAMMs comparisons, Madison

This appendix contains the full set of GAMMs comparisons for each durational scaling /alignment model (see §4.4.3) across all vowels in the Madison dataset. The top set of three plots (in colour) display smooths with confidence intervals for each coda voicing condition for each model; the lower four plots are difference smooths, with the reference level (voiced coda) at top right, followed by the time-normalized model at top left, and the proportionally-scaled duration models, left- and right-alignment, in the lower left and right plots, respectively.

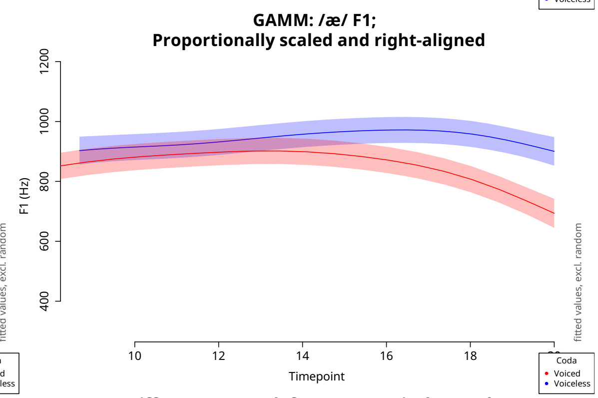
æ F1



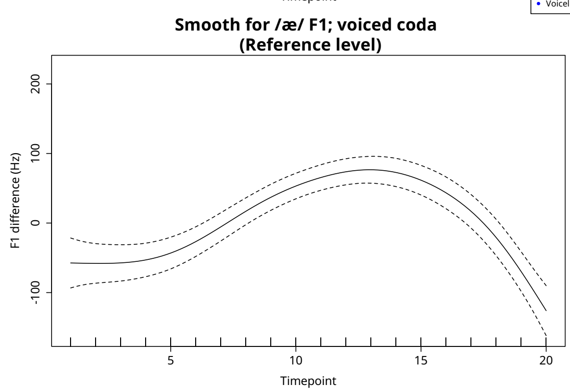
GAMM: /æ/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless



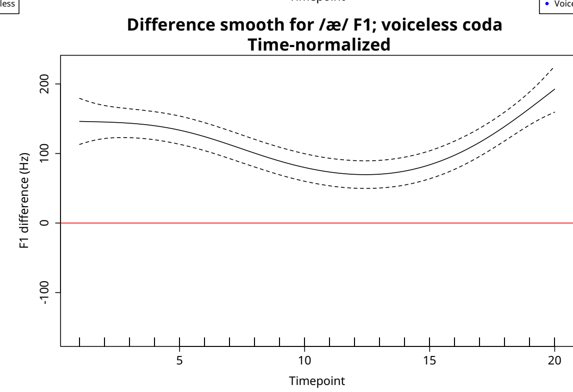
GAMM: /æ/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless



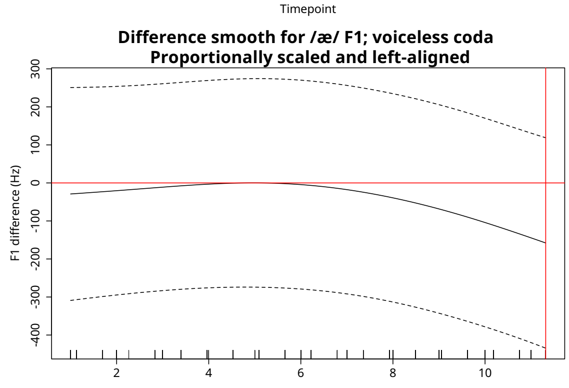
GAMM: /æ/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless



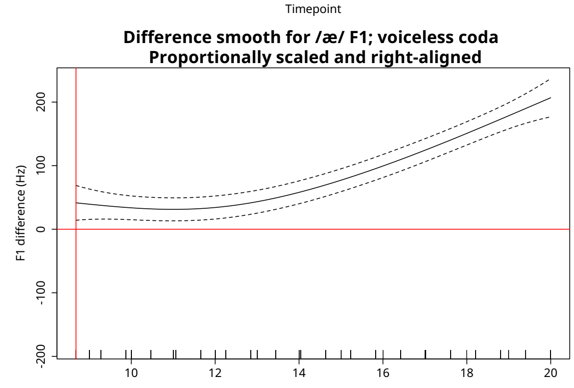
Smooth for /æ/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint



Difference smooth for /æ/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint

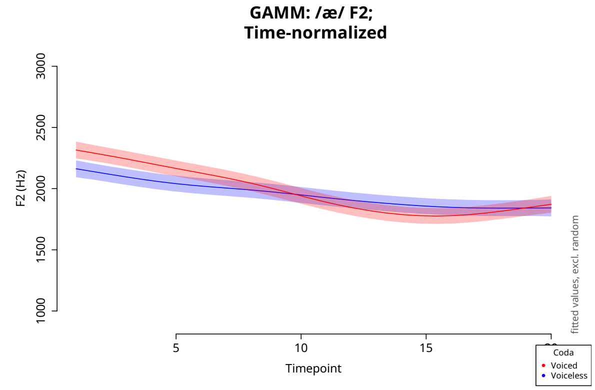


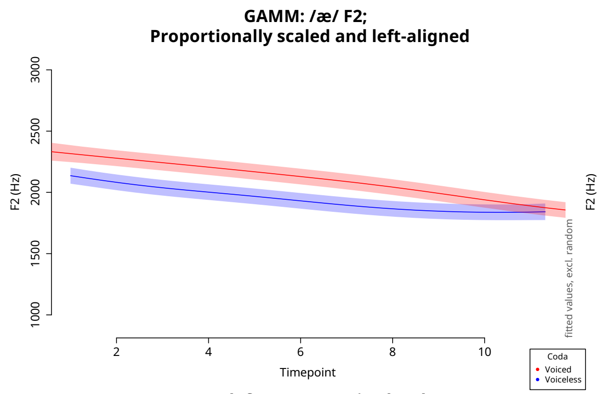
Difference smooth for /æ/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

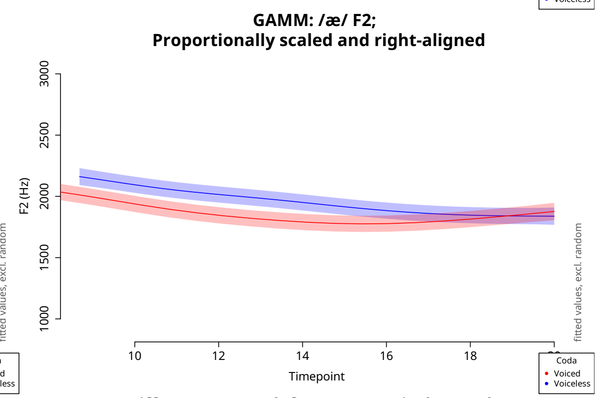


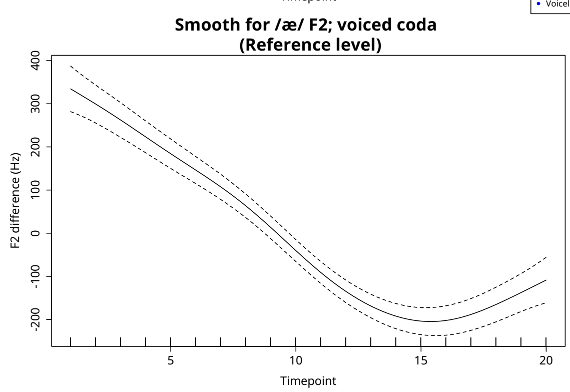
Difference smooth for /æ/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

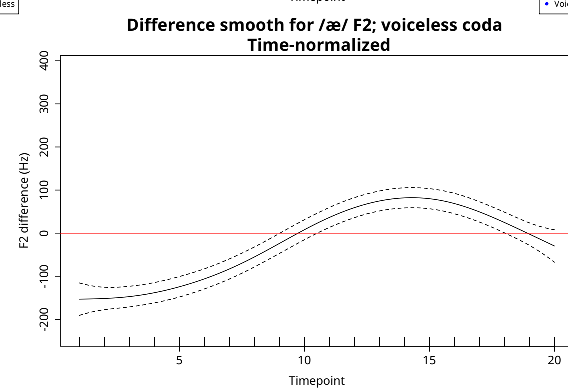
æ F2

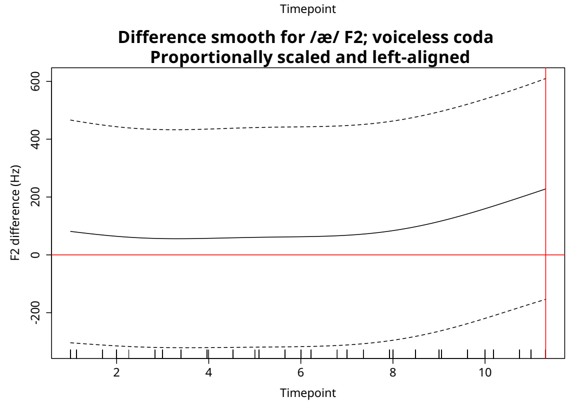

GAMM: /æ/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

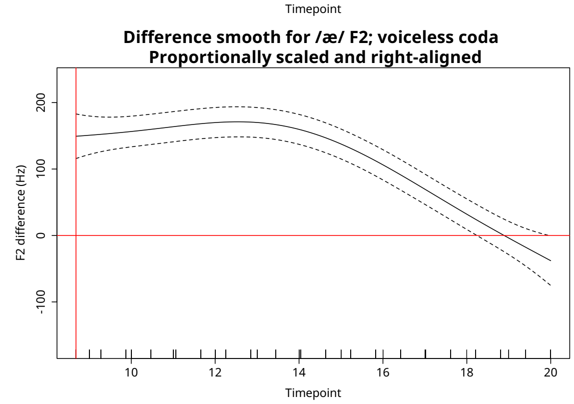

GAMM: /æ/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /æ/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

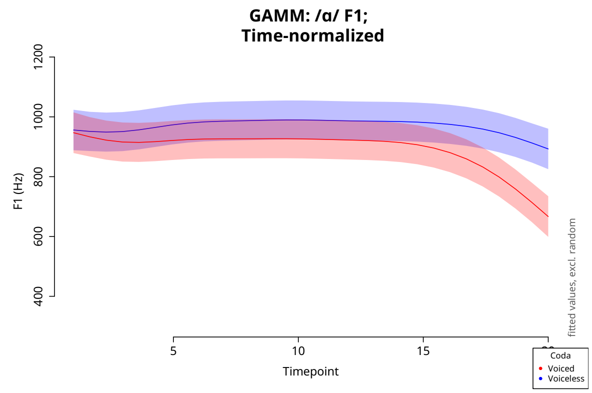

Smooth for /æ/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint

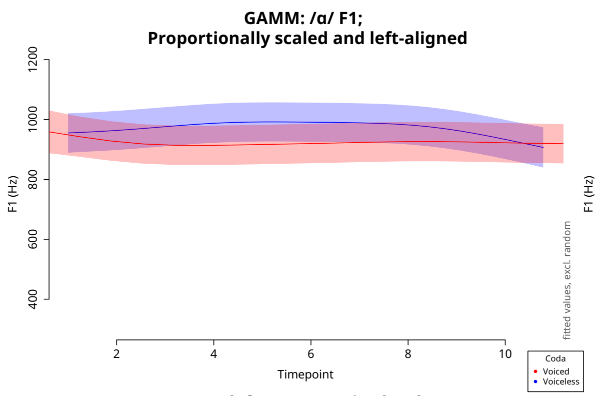

Difference smooth for /æ/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint

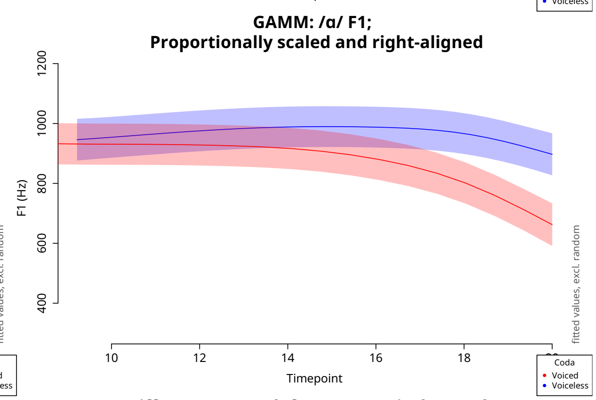

Difference smooth for /æ/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

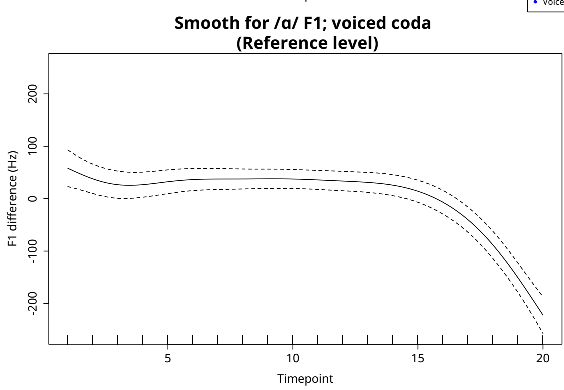

Difference smooth for /æ/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

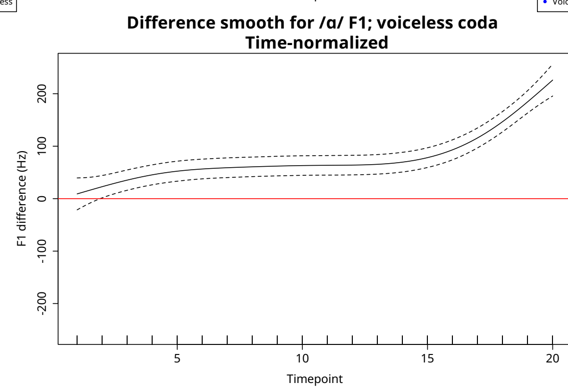
ɑ F1

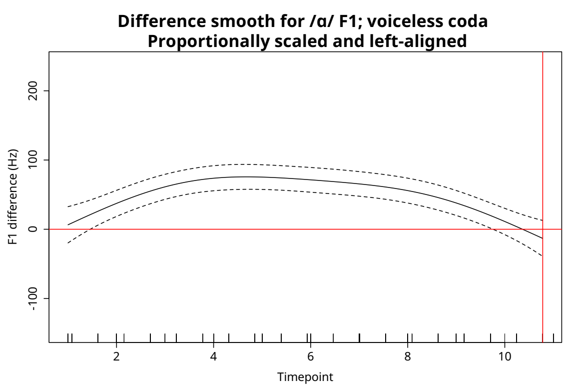

GAMM: /ɑ/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

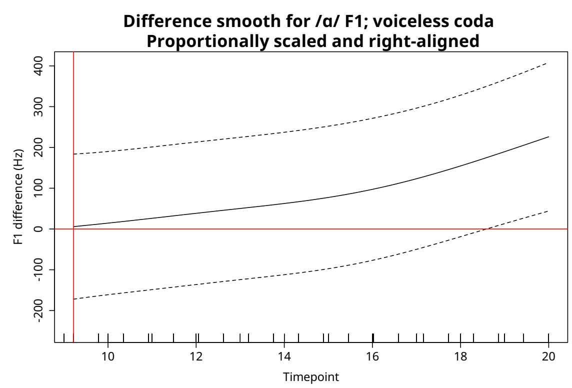

GAMM: /ɑ/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ɑ/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

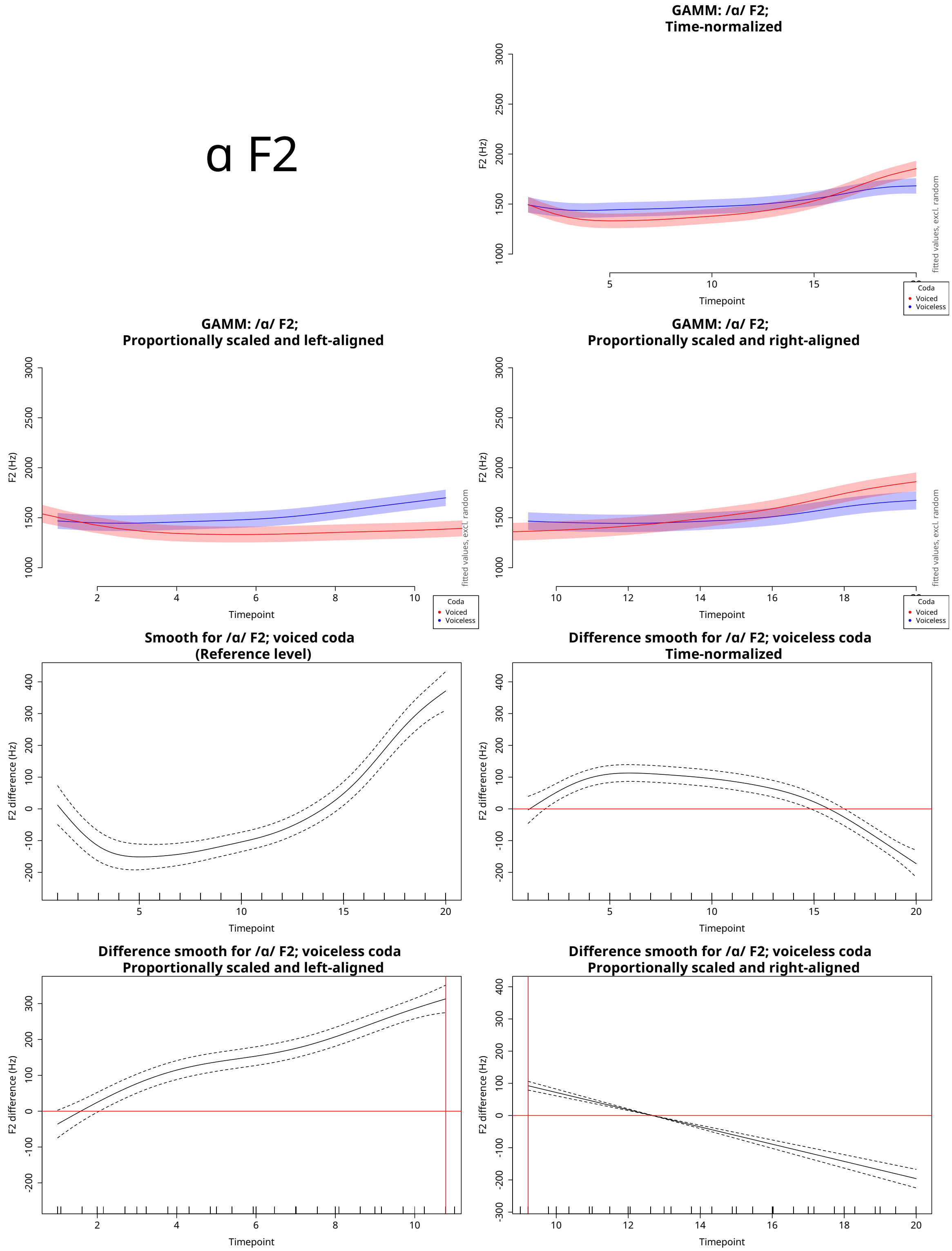

Smooth for /ɑ/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint


Difference smooth for /ɑ/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint


Difference smooth for /ɑ/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

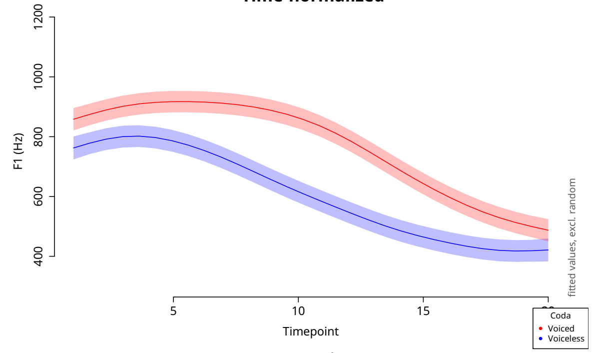

Difference smooth for /ɑ/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

ɑ F2

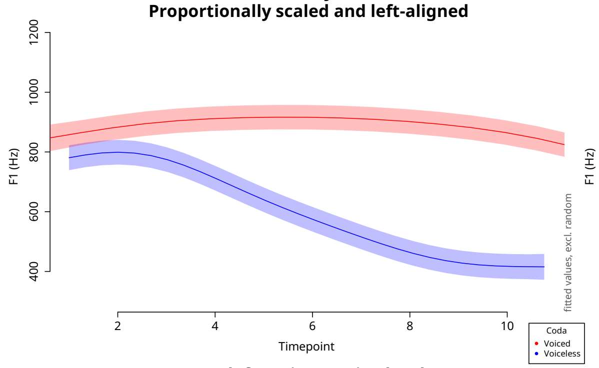

GAMM: /ɑ/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless
GAMM: /ɑ/ F2;
Proportionally scaled and left-aligned
GAMM: /ɑ/ F2;
Proportionally scaled and right-aligned
Smooth for /ɑ/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Difference smooth for /ɑ/ F2; voiceless coda
Time-normalized
Difference smooth for /ɑ/ F2; voiceless coda
Proportionally scaled and left-aligned
Difference smooth for /ɑ/ F2; voiceless coda
Proportionally scaled and right-aligned

aj F1

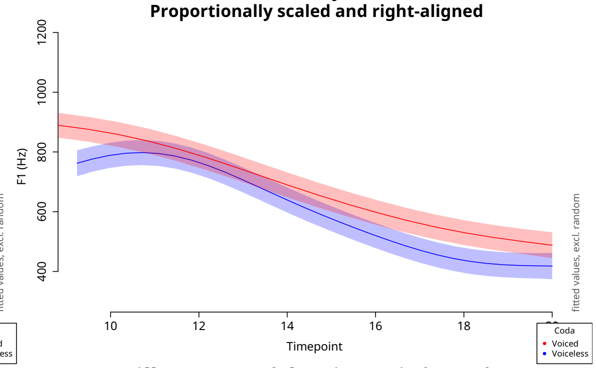
GAMM: /aj/ F1; Time-normalized



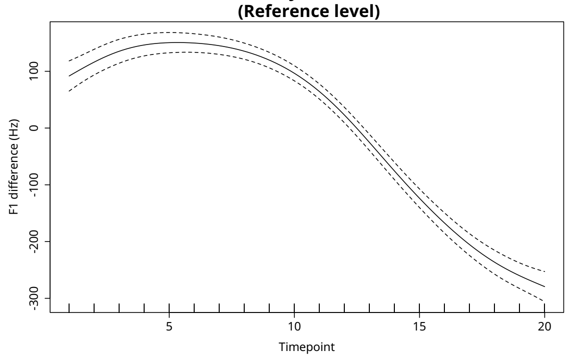
GAMM: /aj/ F1; Proportionally scaled and left-aligned



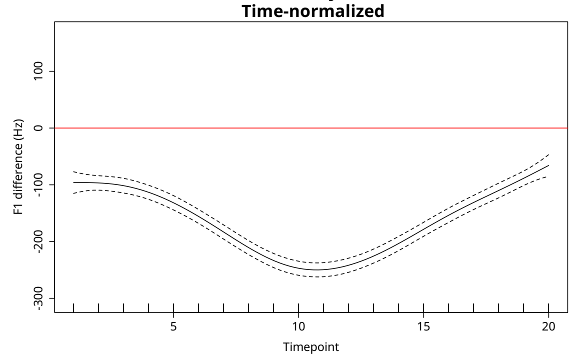
GAMM: /aj/ F1; Proportionally scaled and right-aligned



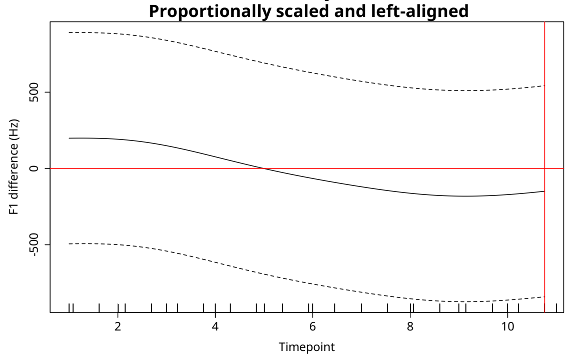
Smooth for /aj/ F1; voiced coda (Reference level)



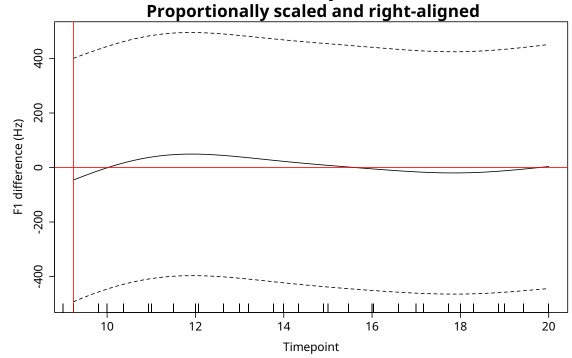
Difference smooth for /aj/ F1; voiceless coda Time-normalized



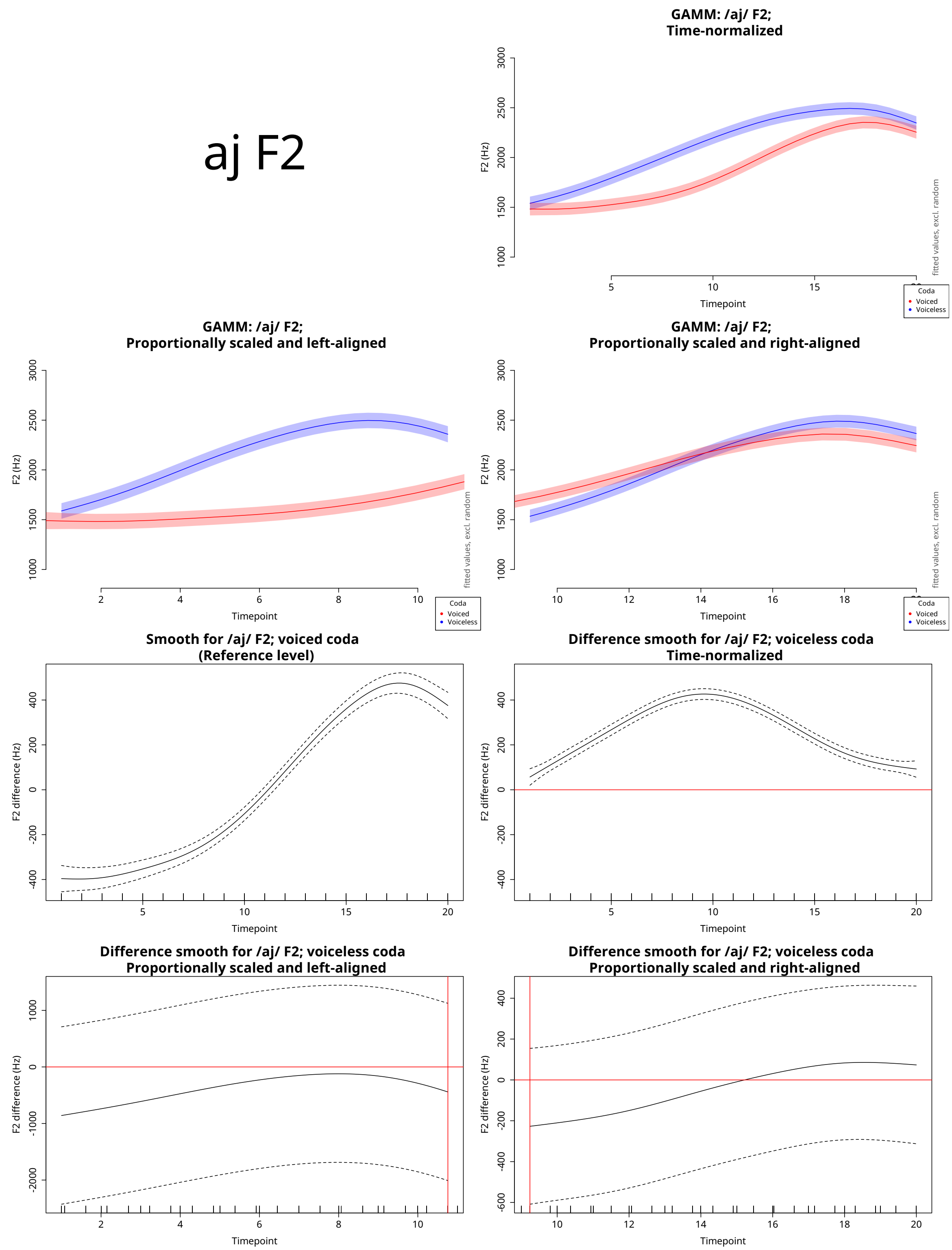
Difference smooth for /aj/ F1; voiceless coda Proportionally scaled and left-aligned



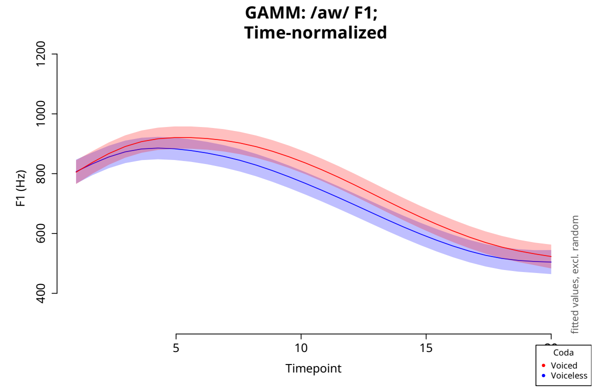
Difference smooth for /aj/ F1; voiceless coda Proportionally scaled and right-aligned

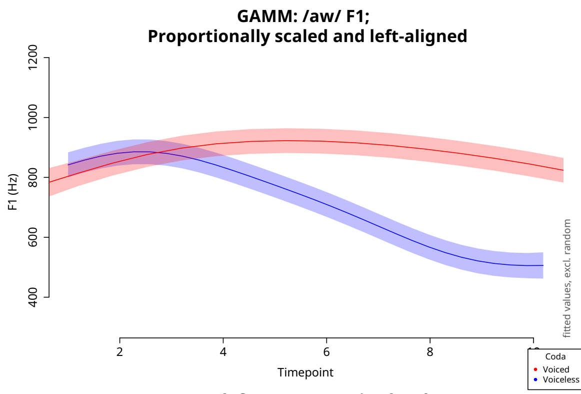


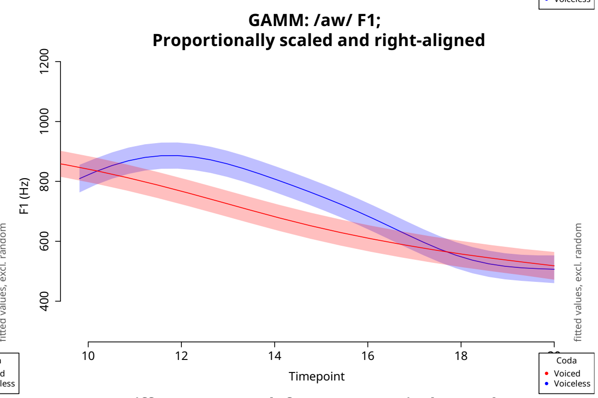
aj F2

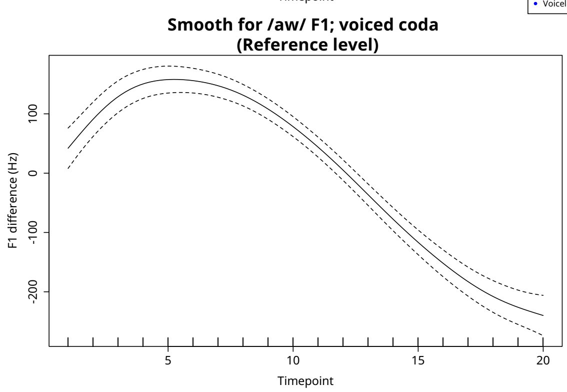

GAMM: /aj/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless
GAMM: /aj/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless
GAMM: /aj/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless
Smooth for /aj/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint
Difference smooth for /aj/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint
Difference smooth for /aj/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint
Difference smooth for /aj/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

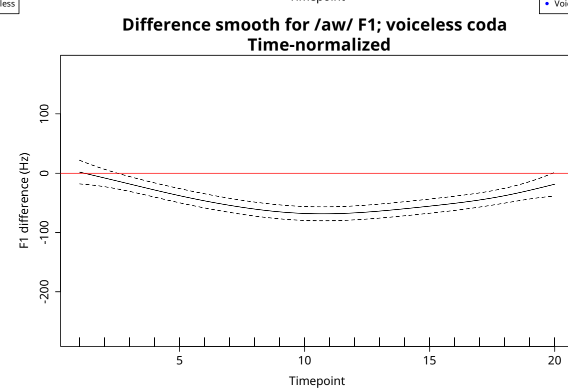
aw F1

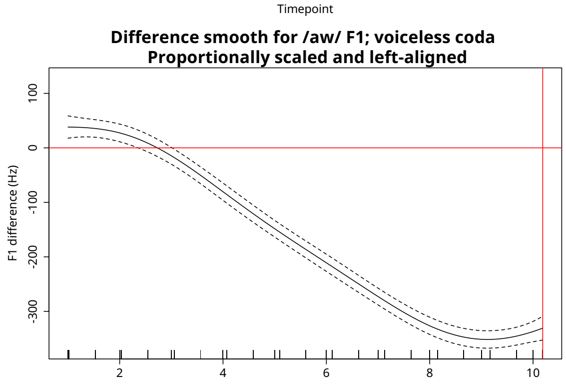

GAMM: /aw/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

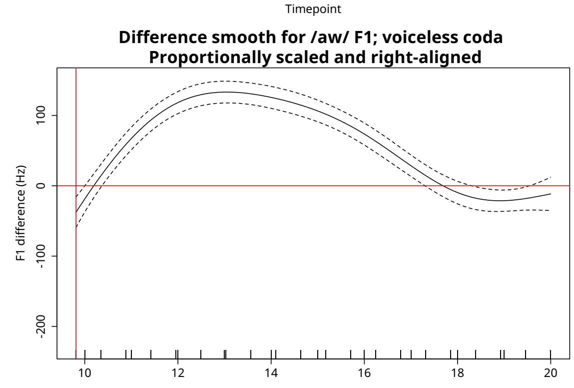

GAMM: /aw/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /aw/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

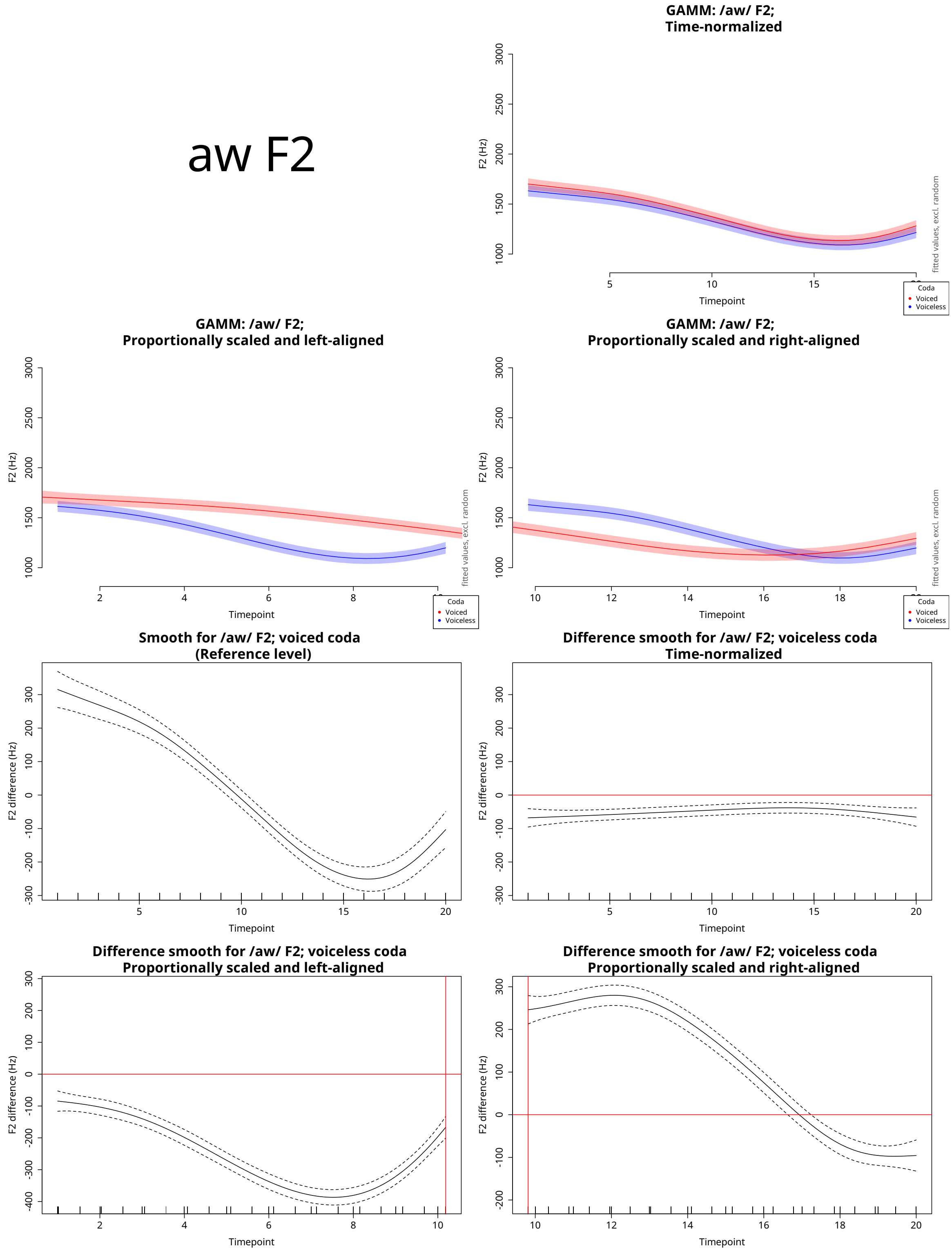

Smooth for /aw/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint


Difference smooth for /aw/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint


Difference smooth for /aw/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

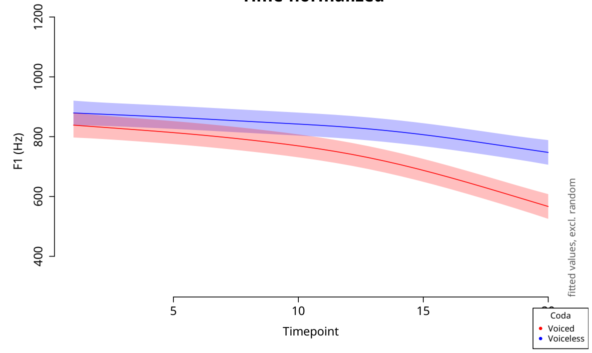

Difference smooth for /aw/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

aw F2

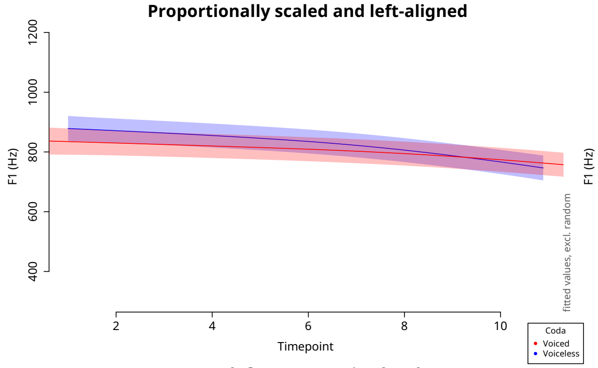

GAMM: /aw/ F2;
Time-normalized
GAMM: /aw/ F2;
Proportionally scaled and left-aligned
GAMM: /aw/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless
Smooth for /aw/ F2; voiced coda
(Reference level)
Difference smooth for /aw/ F2; voiceless coda
Time-normalized
Difference smooth for /aw/ F2; voiceless coda
Proportionally scaled and left-aligned
Difference smooth for /aw/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)

ɛ F1

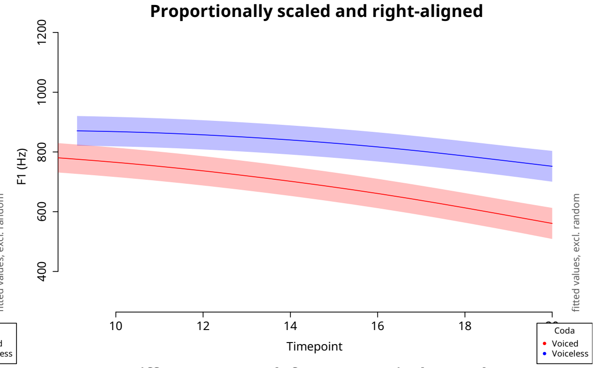
GAMM: /ɛ/ F1;
Time-normalized



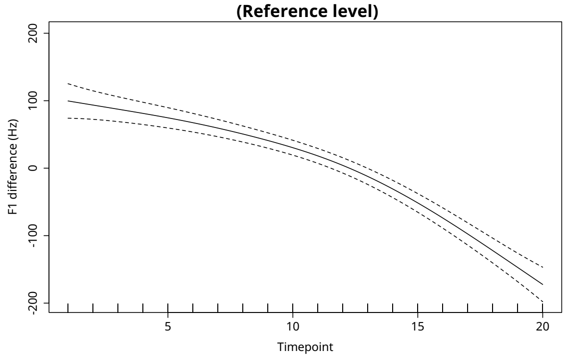
GAMM: /ɛ/ F1;
Proportionally scaled and left-aligned



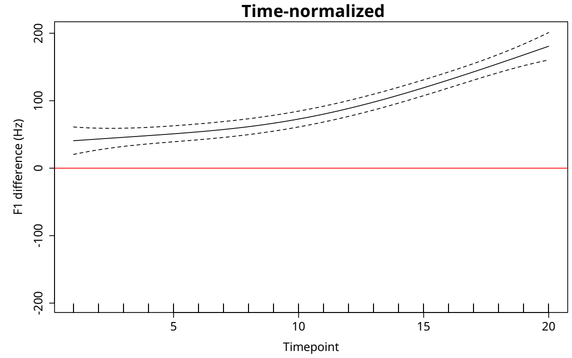
GAMM: /ɛ/ F1;
Proportionally scaled and right-aligned



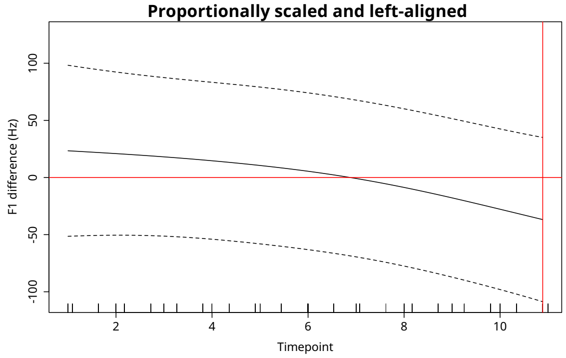
Smooth for /ɛ/ F1; voiced coda
(Reference level)



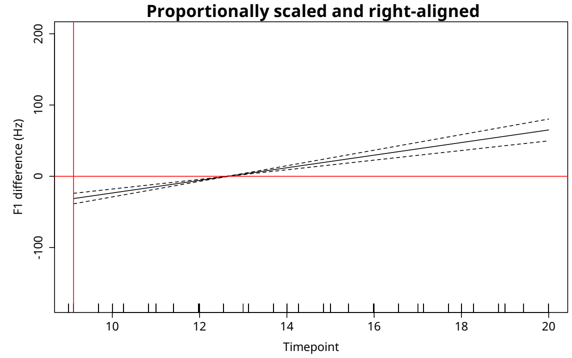
Difference smooth for /ɛ/ F1; voiceless coda
Time-normalized



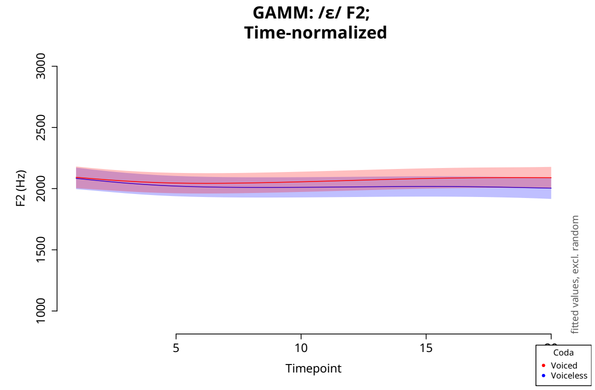
Difference smooth for /ɛ/ F1; voiceless coda
Proportionally scaled and left-aligned

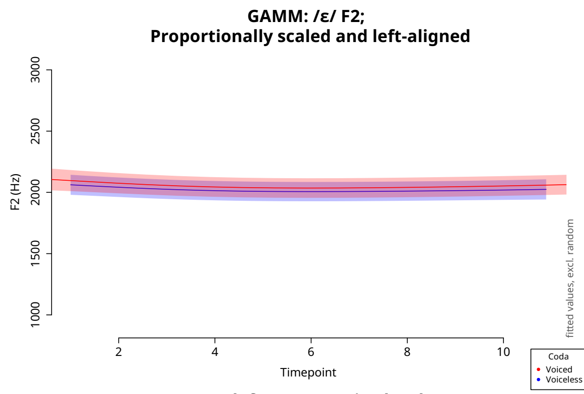


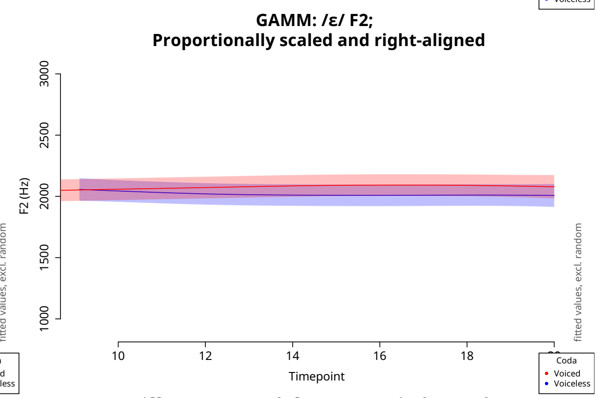
Difference smooth for /ɛ/ F1; voiceless coda
Proportionally scaled and right-aligned

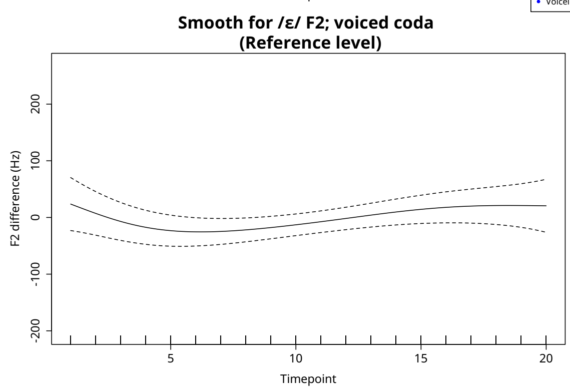


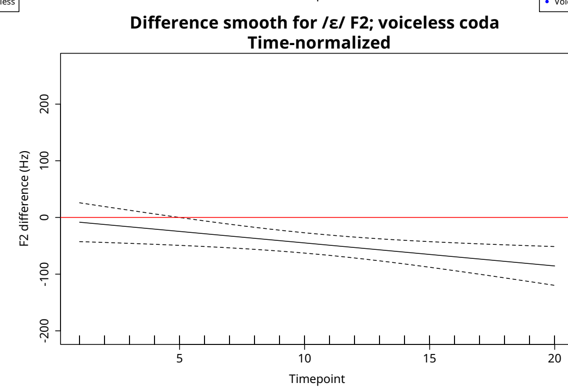
ɛ F2

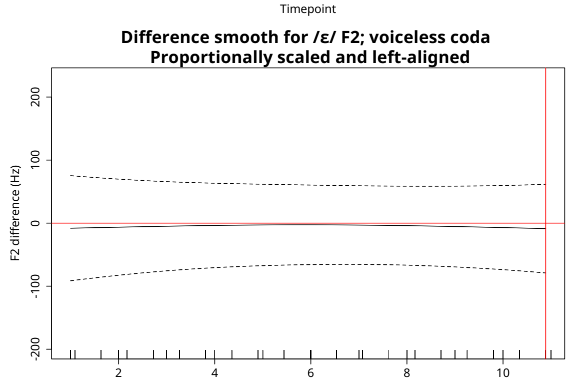

GAMM: /ɛ/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

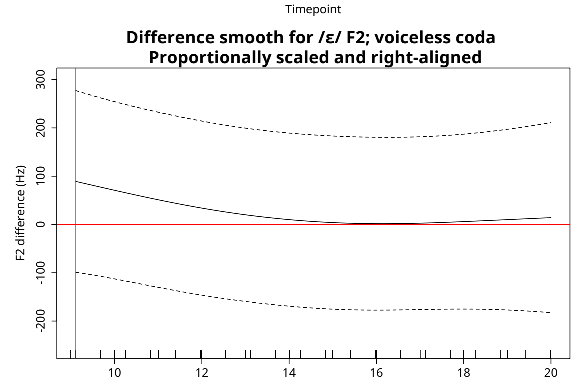

GAMM: /ɛ/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ɛ/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

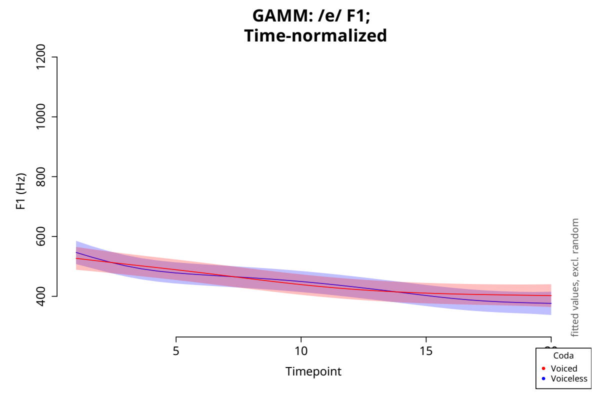

Smooth for /ɛ/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint


Difference smooth for /ɛ/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint

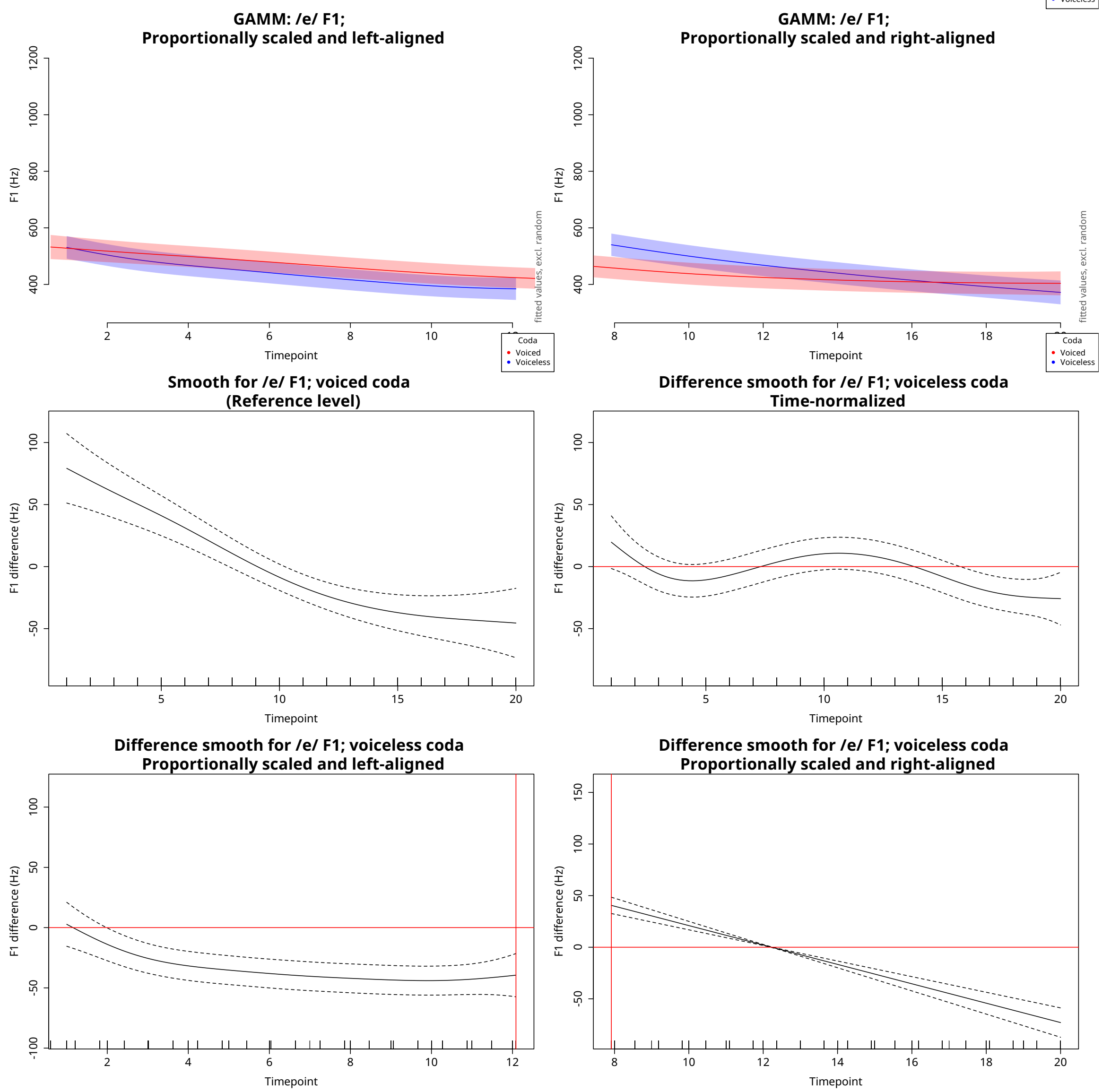

Difference smooth for /ɛ/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint


Difference smooth for /ɛ/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

e F1

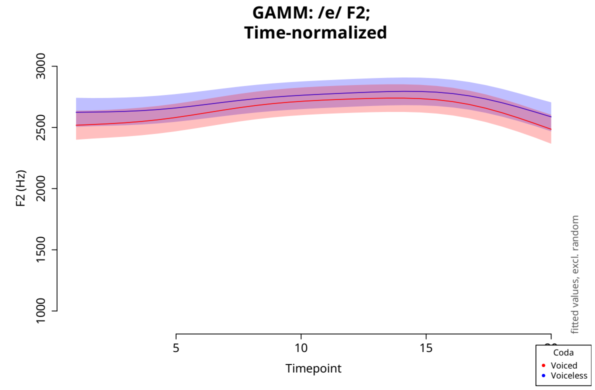


GAMM: /e/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

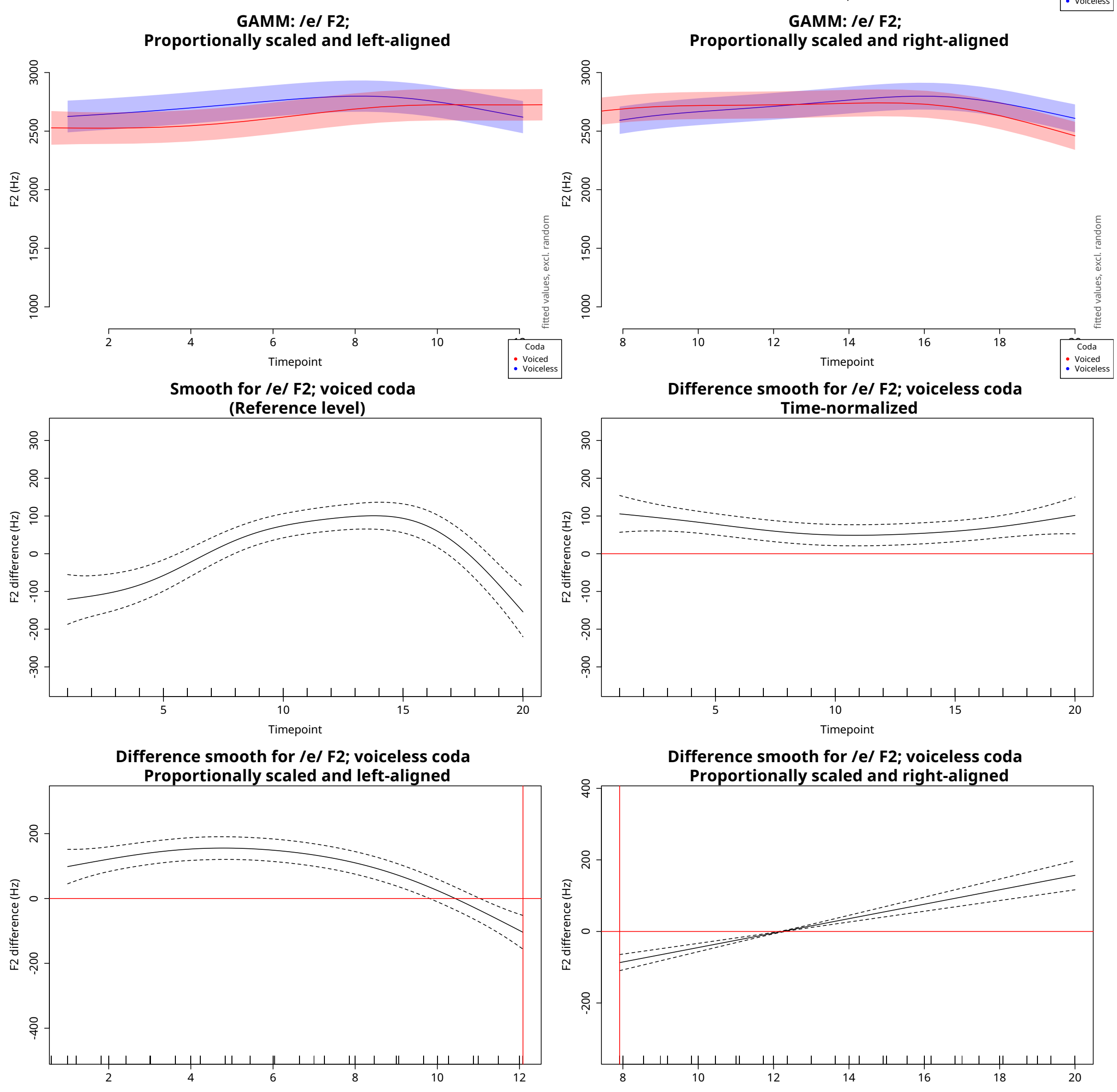


GAMM: /e/ F1;
Proportionally scaled and left-aligned
GAMM: /e/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless
Smooth for /e/ F1; voiced coda
(Reference level)
Difference smooth for /e/ F1; voiceless coda
Time-normalized
Difference smooth for /e/ F1; voiceless coda
Proportionally scaled and left-aligned
Difference smooth for /e/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

e F2

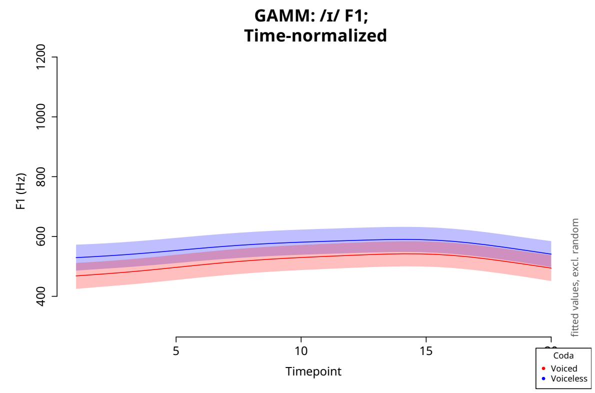


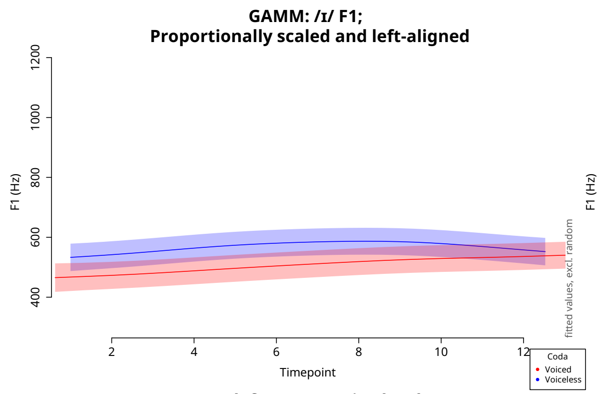
GAMM: /e/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

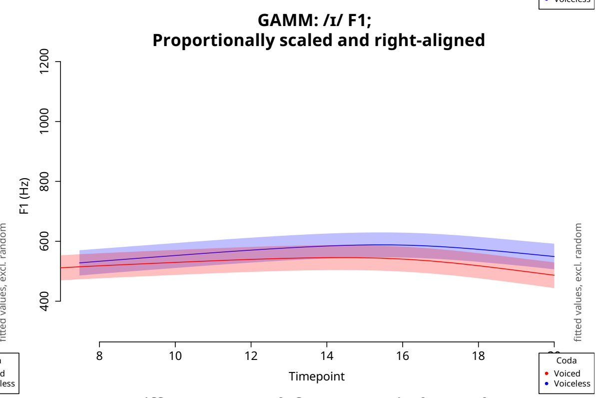


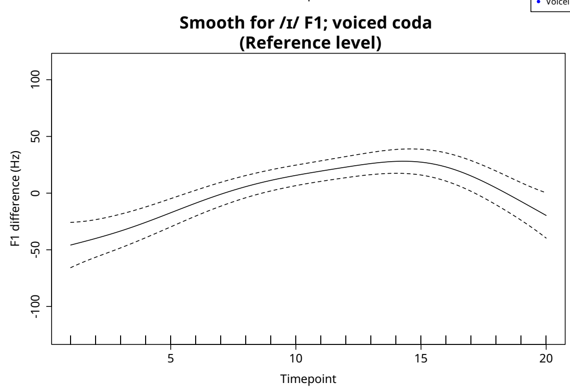
GAMM: /e/ F2;
Proportionally scaled and left-aligned
GAMM: /e/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless
Smooth for /e/ F2; voiced coda
(Reference level)
Difference smooth for /e/ F2; voiceless coda
Time-normalized
Difference smooth for /e/ F2; voiceless coda
Proportionally scaled and left-aligned
Difference smooth for /e/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

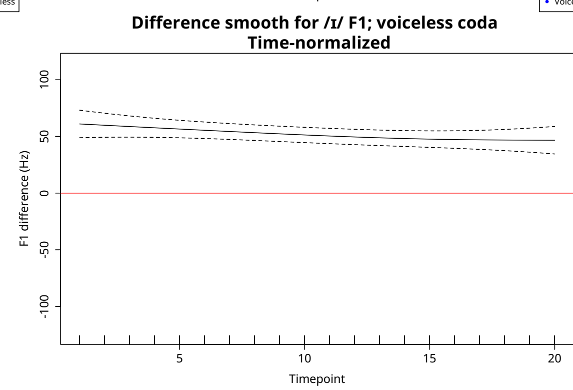
ɪ F1

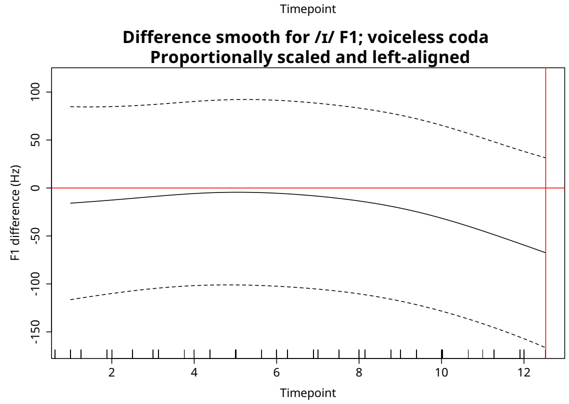

GAMM: /ɪ/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

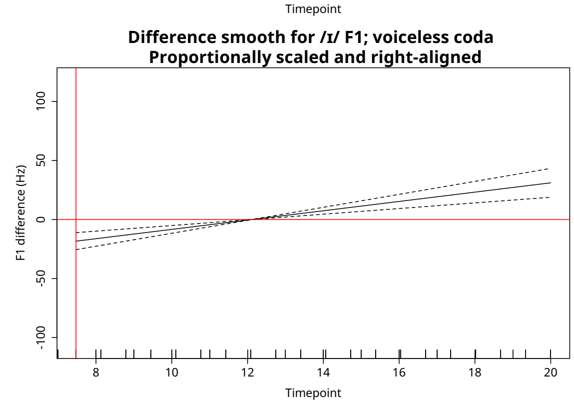

GAMM: /ɪ/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ɪ/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

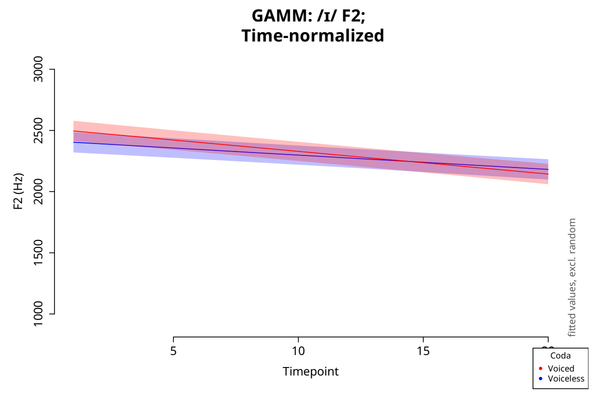

Smooth for /ɪ/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint

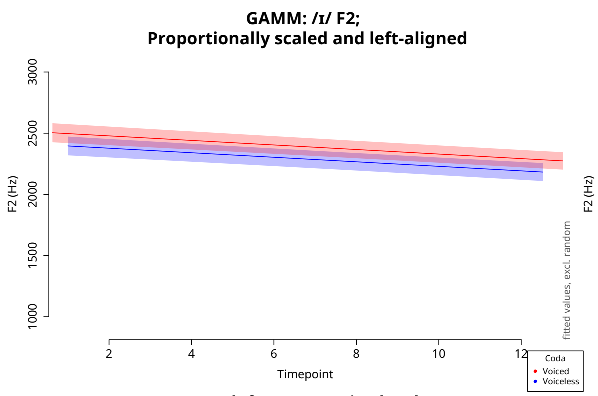

Difference smooth for /ɪ/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint

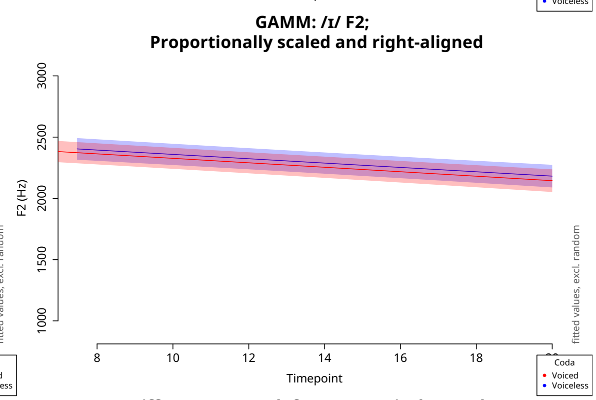

Difference smooth for /ɪ/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

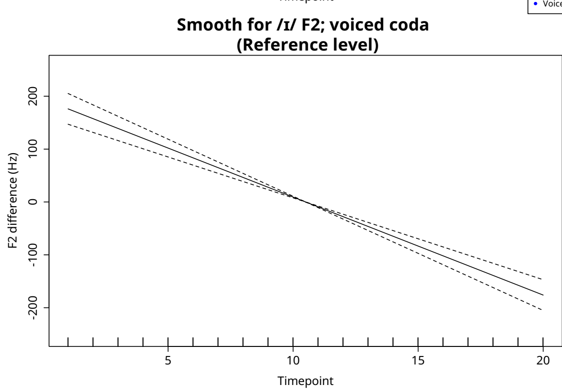

Difference smooth for /ɪ/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

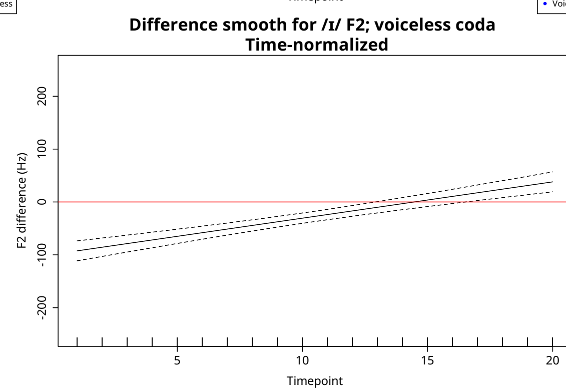
ɪ F2

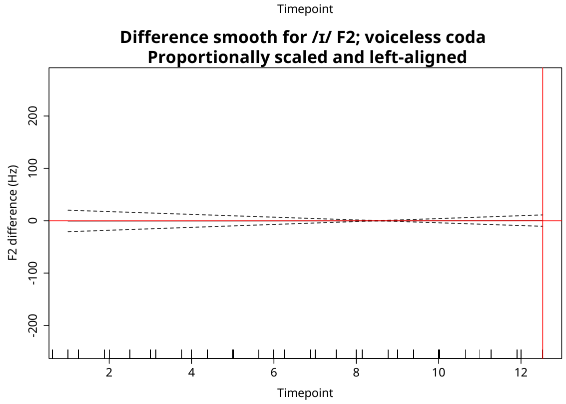

GAMM: /ɪ/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

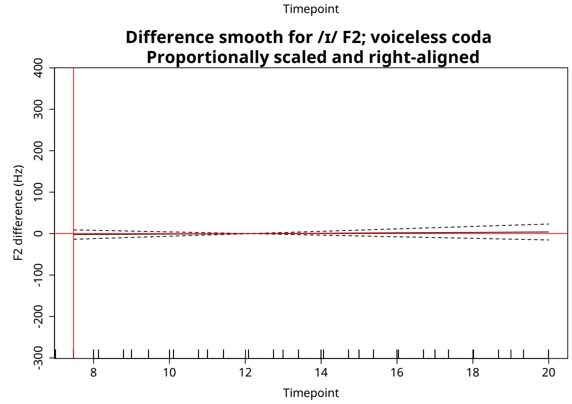

GAMM: /ɪ/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ɪ/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

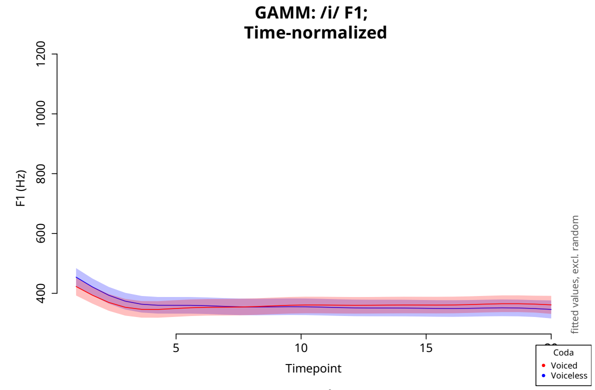

Smooth for /ɪ/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint

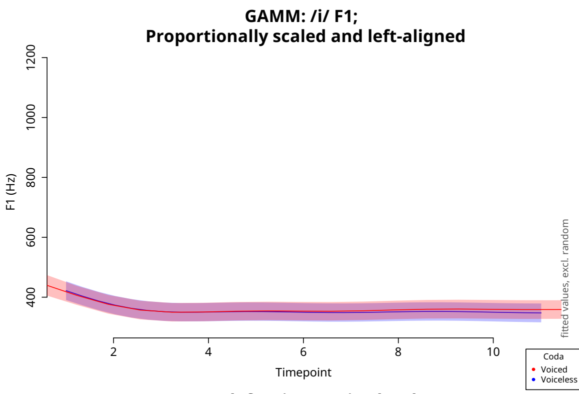

Difference smooth for /ɪ/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint

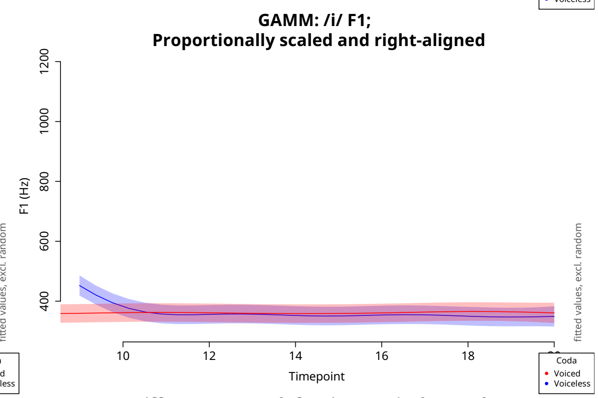

Difference smooth for /ɪ/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

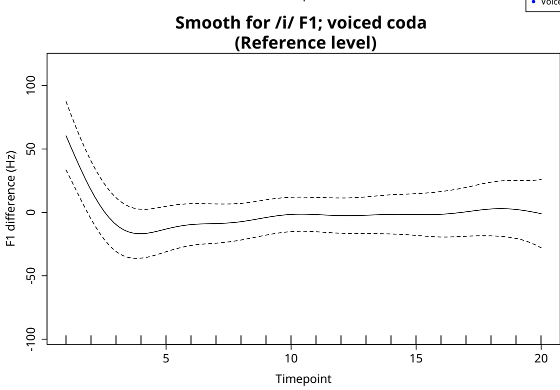

Difference smooth for /ɪ/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

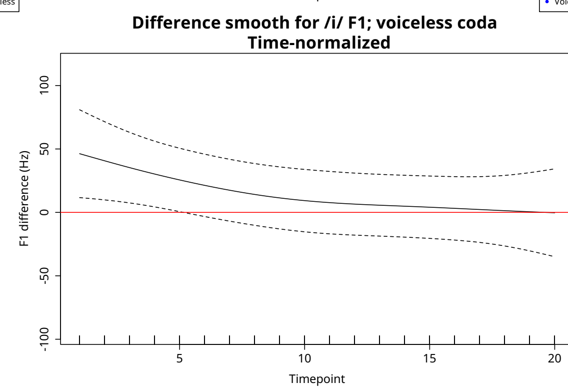
i F1

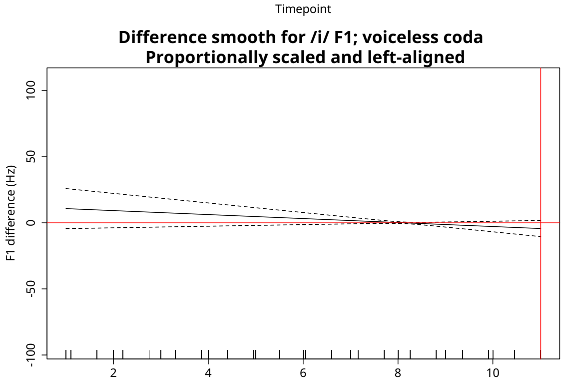

GAMM: /i/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

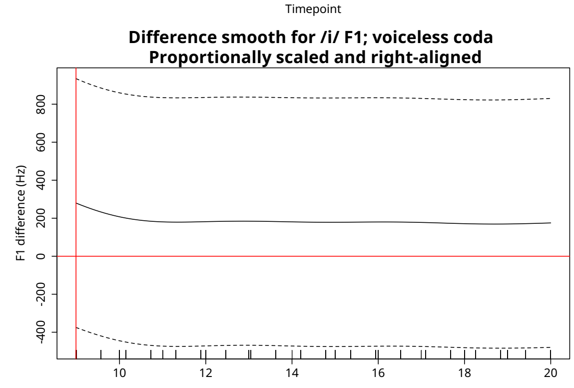

GAMM: /i/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /i/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

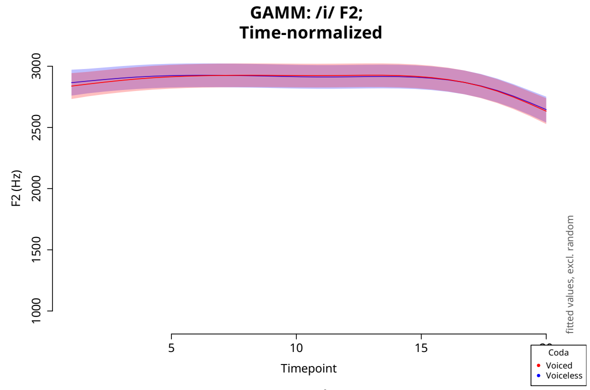

Smooth for /i/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint

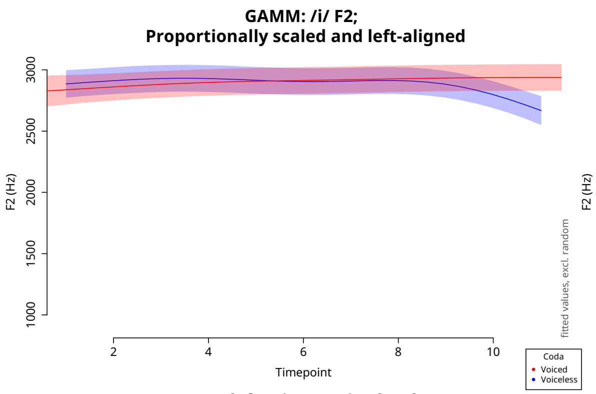

Difference smooth for /i/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint

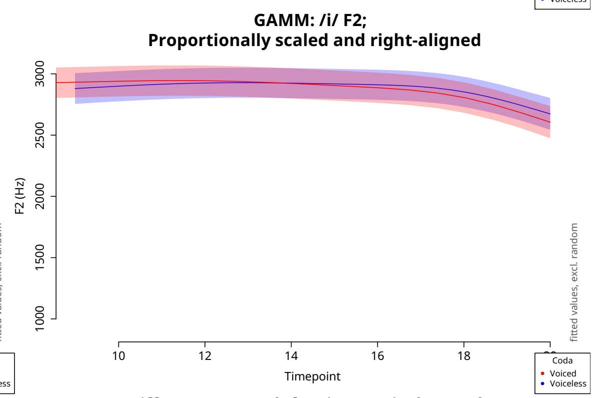

Difference smooth for /i/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

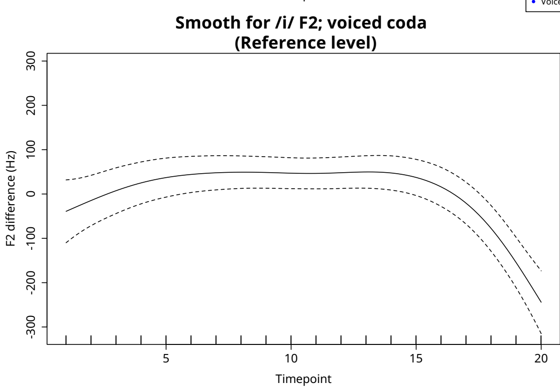

Difference smooth for /i/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

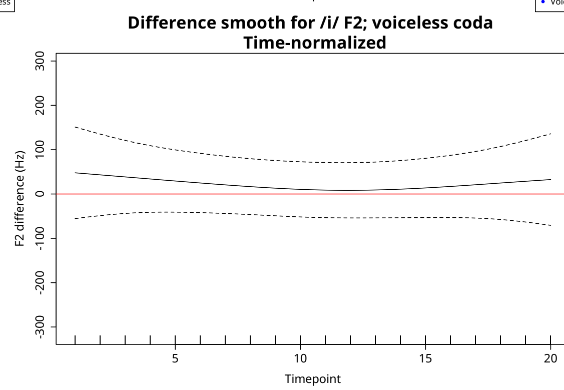
i F2

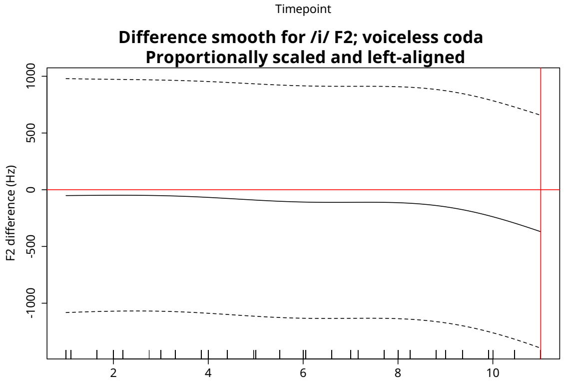

GAMM: /i/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

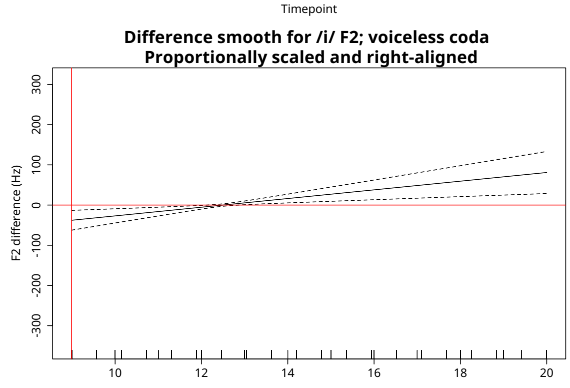

GAMM: /i/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /i/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


Smooth for /i/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint

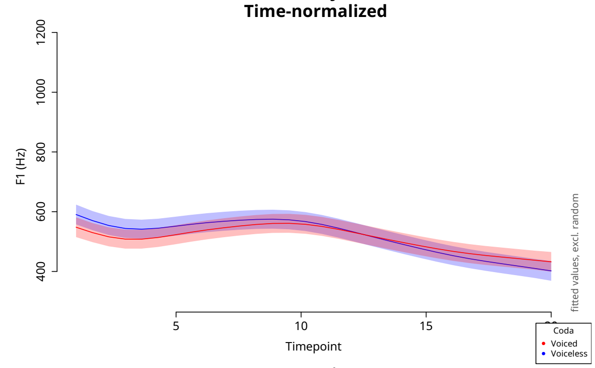

Difference smooth for /i/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint


Difference smooth for /i/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

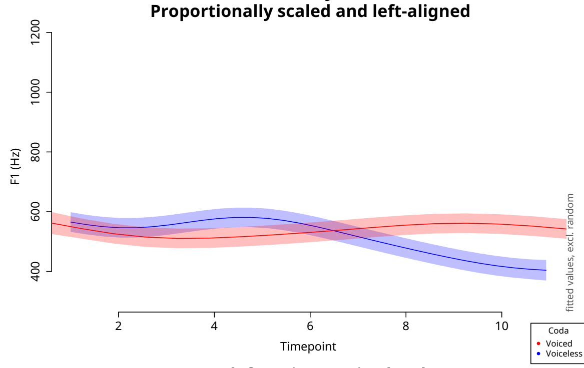

Difference smooth for /i/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

ɔj F1

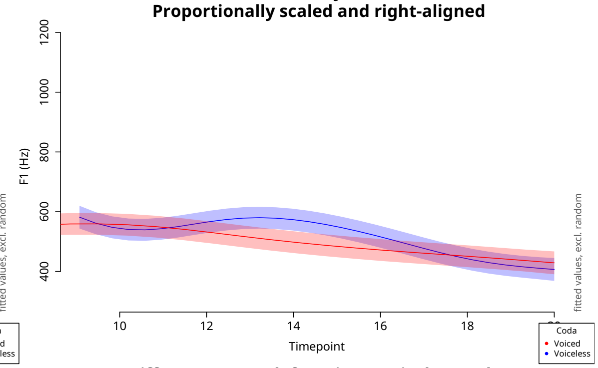
GAMM: /ɔj/ F1;
Time-normalized



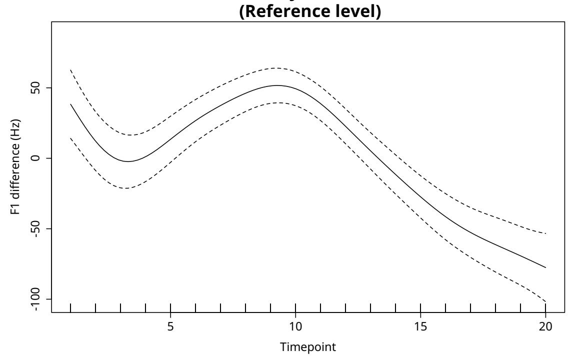
GAMM: /ɔj/ F1;
Proportionally scaled and left-aligned



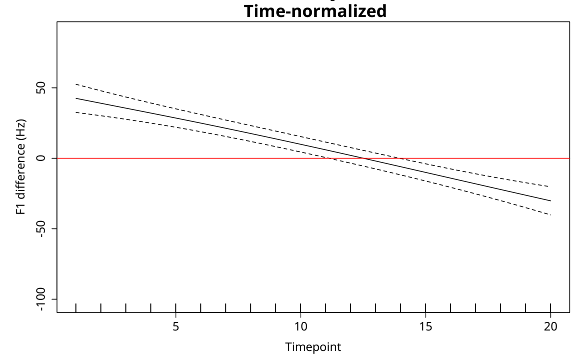
GAMM: /ɔj/ F1;
Proportionally scaled and right-aligned



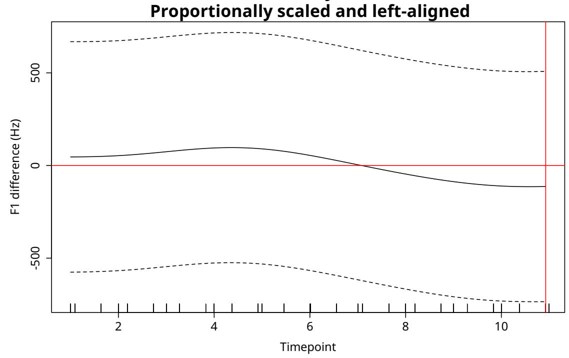
Smooth for /ɔj/ F1; voiced coda
(Reference level)



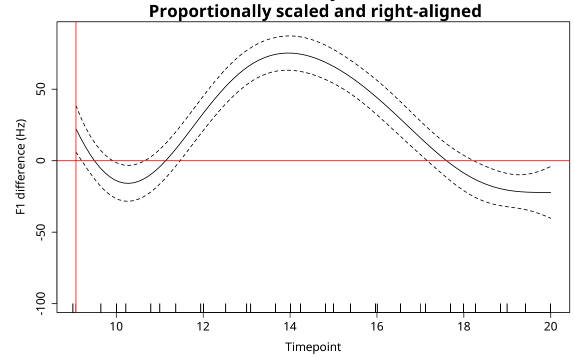
Difference smooth for /ɔj/ F1; voiceless coda
Time-normalized



Difference smooth for /ɔj/ F1; voiceless coda
Proportionally scaled and left-aligned

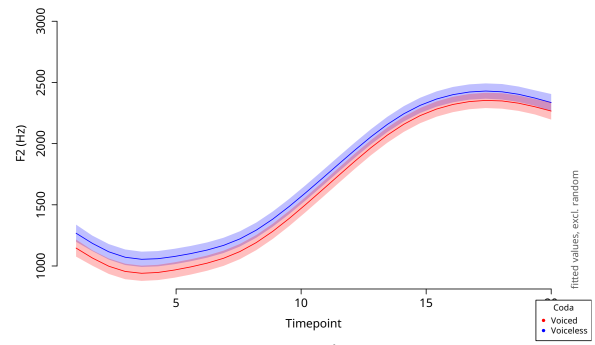


Difference smooth for /ɔj/ F1; voiceless coda
Proportionally scaled and right-aligned

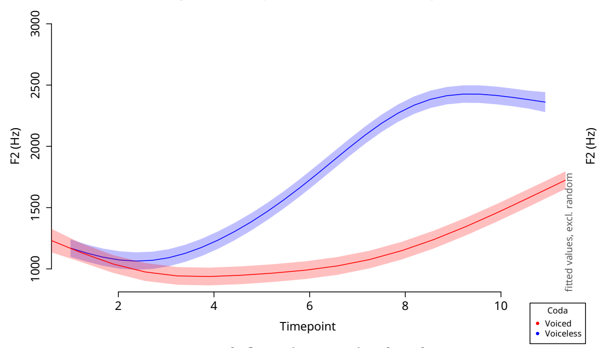


ɔj F2

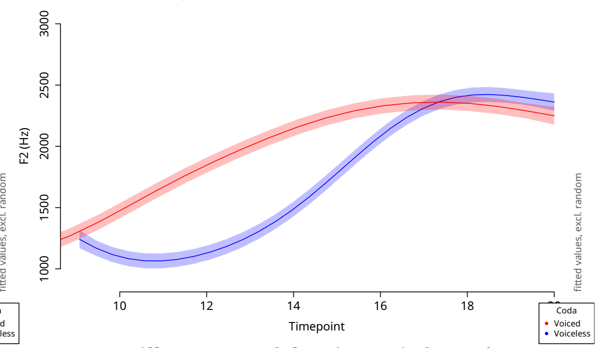
GAMM: /ɔj/ F2;
Time-normalized



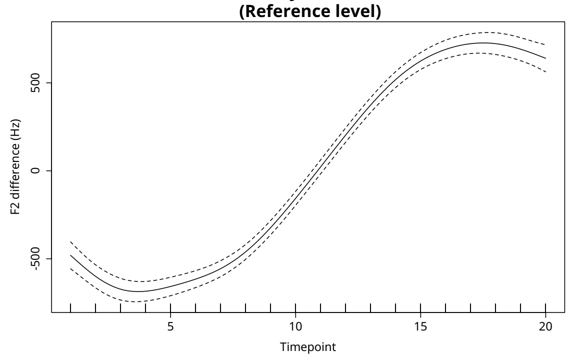
GAMM: /ɔj/ F2;
Proportionally scaled and left-aligned



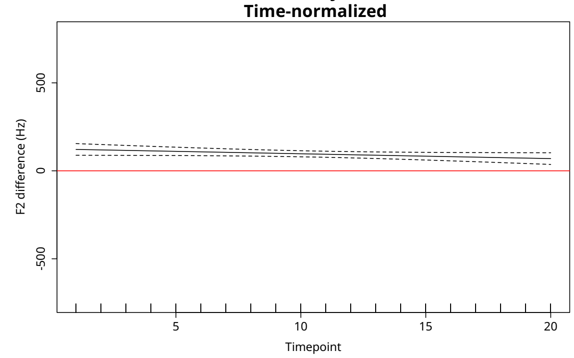
GAMM: /ɔj/ F2;
Proportionally scaled and right-aligned



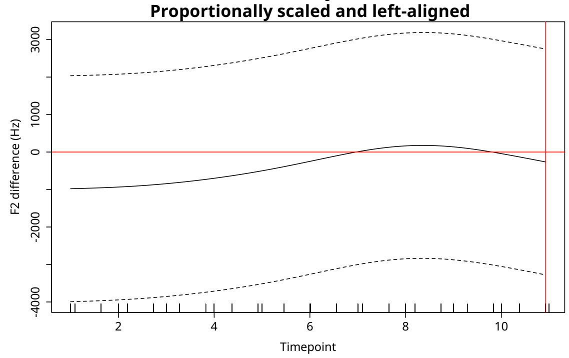
Smooth for /ɔj/ F2; voiced coda
(Reference level)



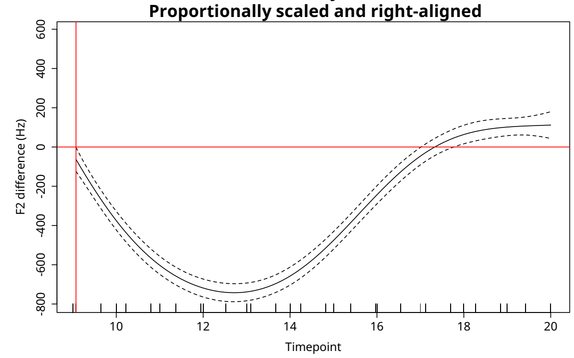
Difference smooth for /ɔj/ F2; voiceless coda
Time-normalized



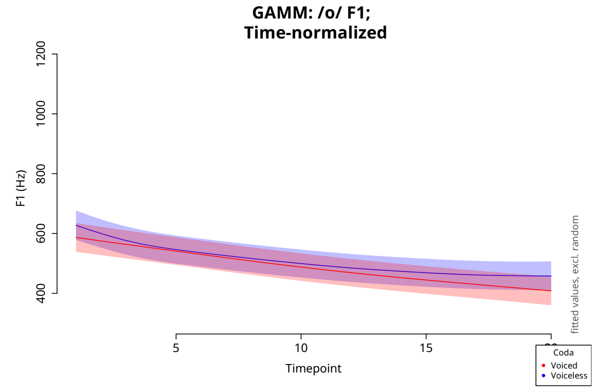
Difference smooth for /ɔj/ F2; voiceless coda
Proportionally scaled and left-aligned



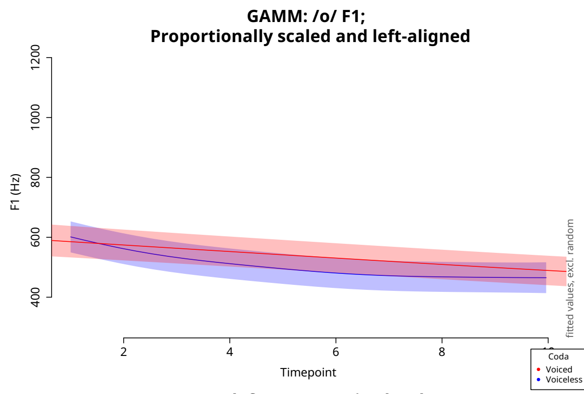
Difference smooth for /ɔj/ F2; voiceless coda
Proportionally scaled and right-aligned



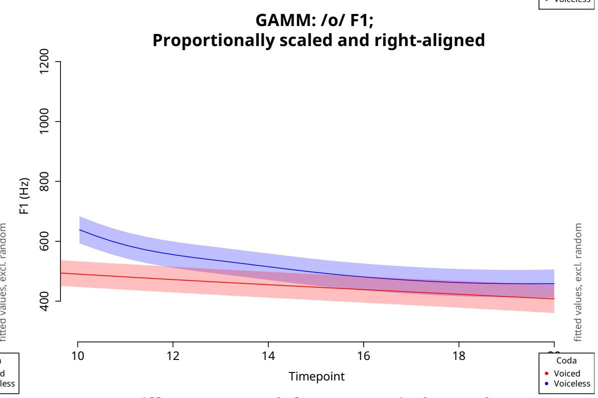
o F1



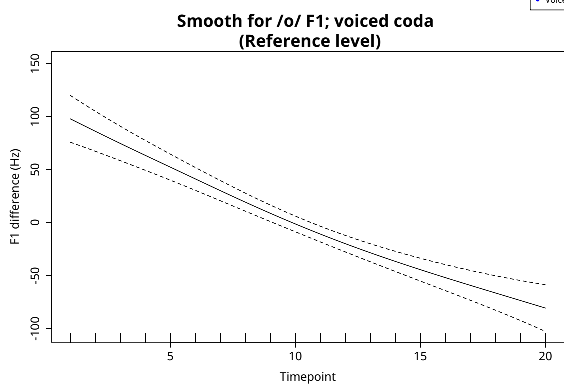
GAMM: /o/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless



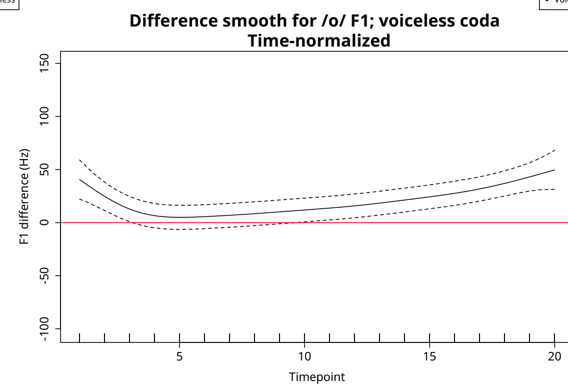
GAMM: /o/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless



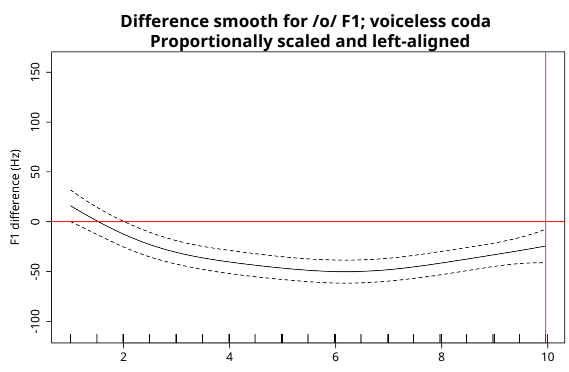
GAMM: /o/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless



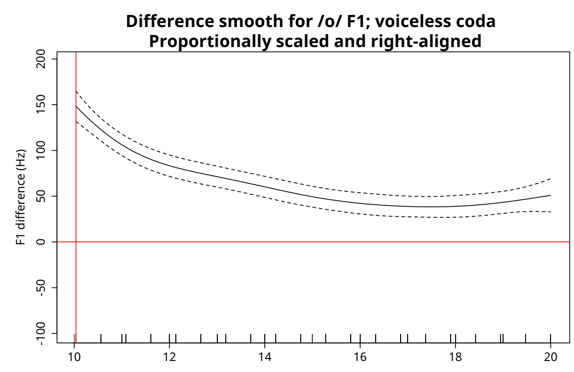
Smooth for /o/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint



Difference smooth for /o/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint

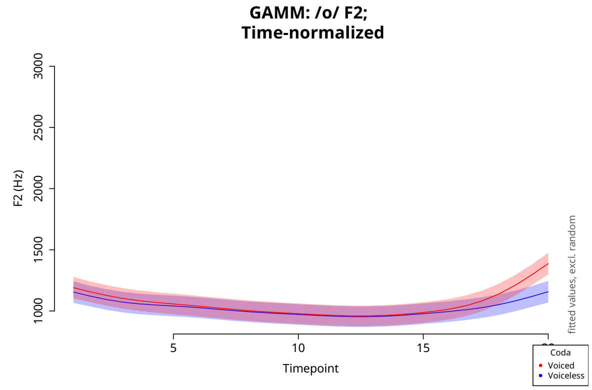


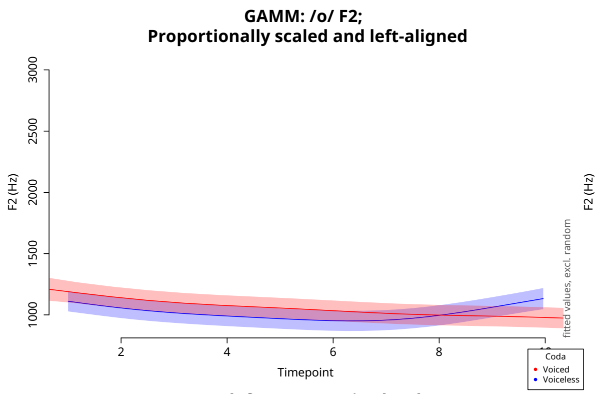
Difference smooth for /o/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

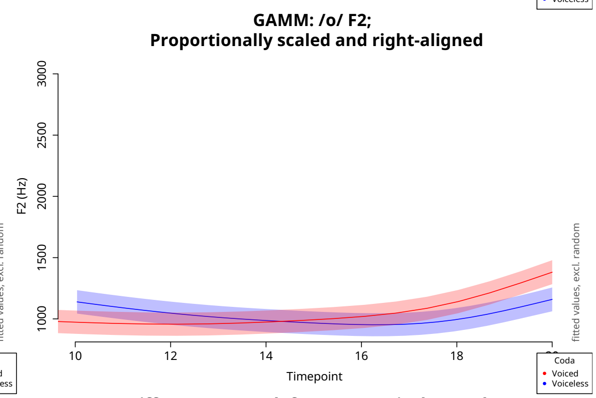


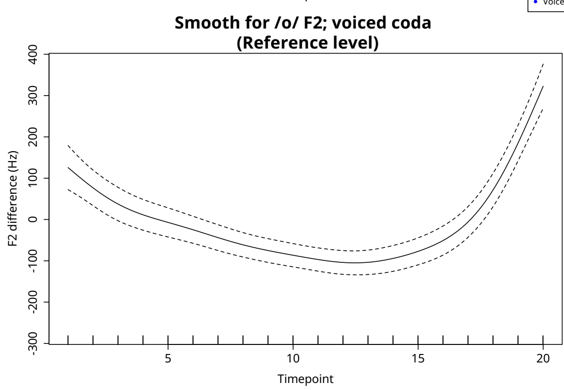
Difference smooth for /o/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

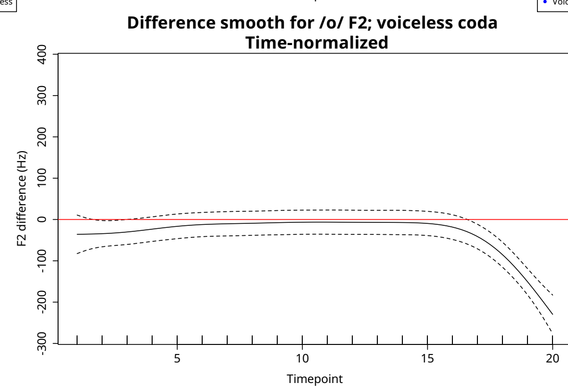
o F2

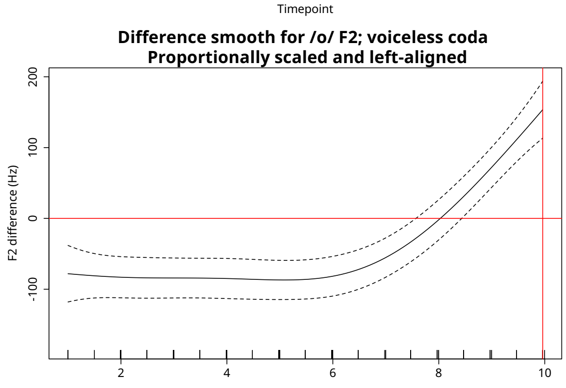

GAMM: /o/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

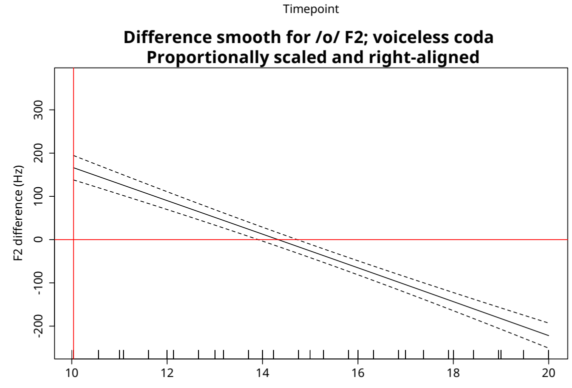

GAMM: /o/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /o/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

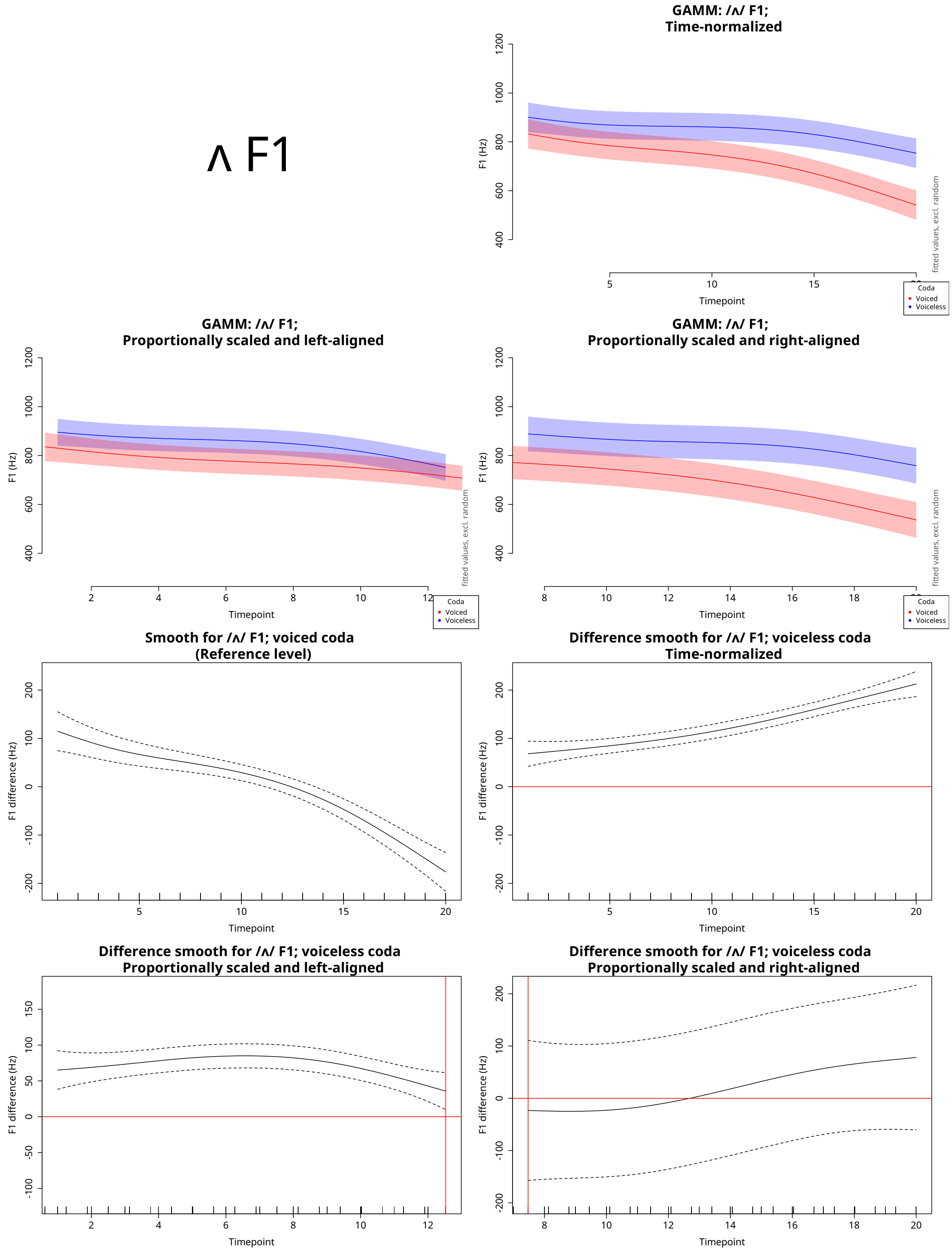

Smooth for /o/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint


Difference smooth for /o/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint

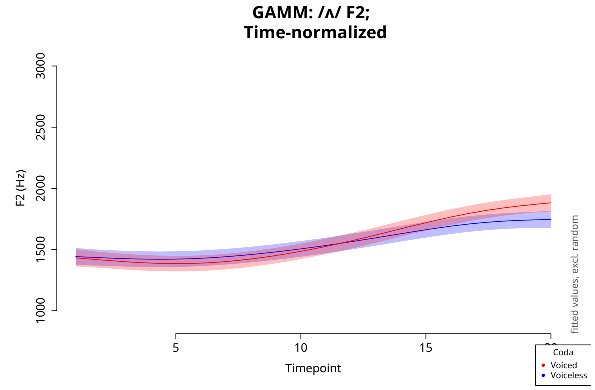

Difference smooth for /o/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

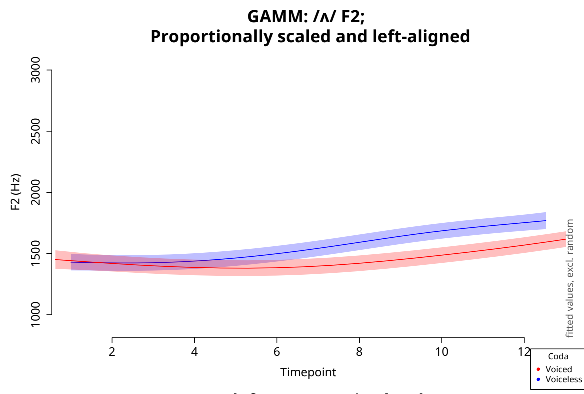

Difference smooth for /o/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

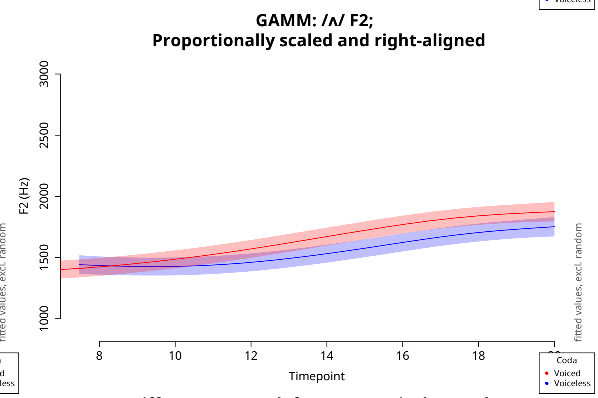
ʌ F1

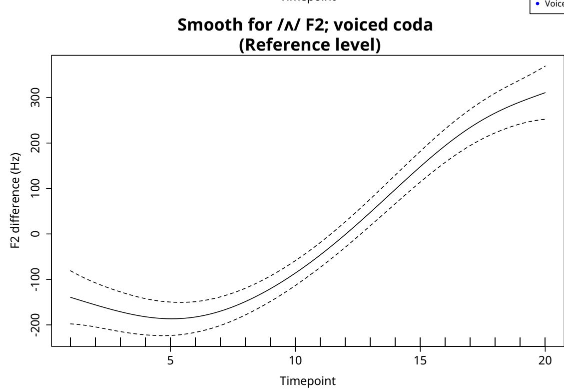

GAMM: /ʌ/ F1;
Time-normalized
GAMM: /ʌ/ F1;
Proportionally scaled and left-aligned
GAMM: /ʌ/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless
Smooth for /ʌ/ F1; voiced coda
(Reference level)
Difference smooth for /ʌ/ F1; voiceless coda
Time-normalized
Difference smooth for /ʌ/ F1; voiceless coda
Proportionally scaled and left-aligned
Difference smooth for /ʌ/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)

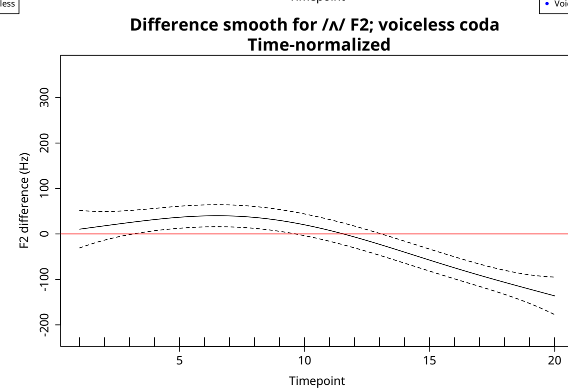
ʌ F2

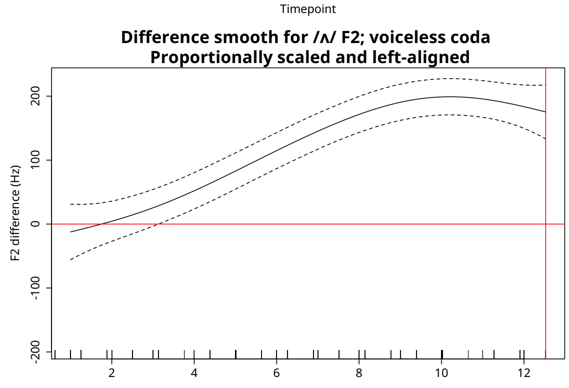

GAMM: /ʌ/ F2;
Time-normalized
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

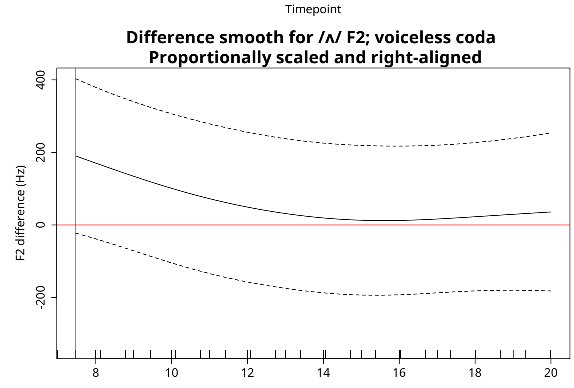

GAMM: /ʌ/ F2;
Proportionally scaled and left-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /ʌ/ F2;
Proportionally scaled and right-aligned
F2 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

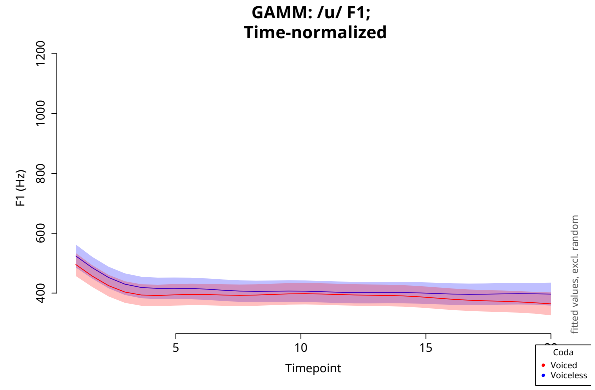

Smooth for /ʌ/ F2; voiced coda
(Reference level)
F2 difference (Hz)
Timepoint

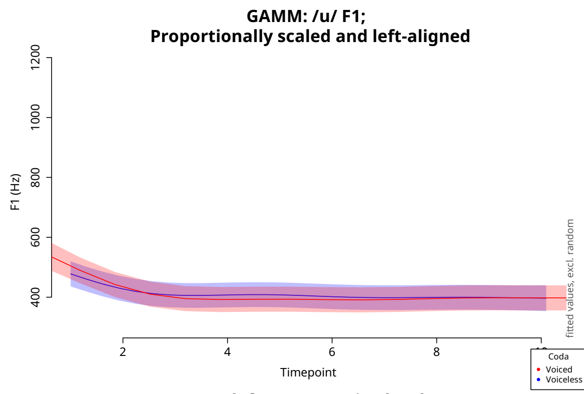

Difference smooth for /ʌ/ F2; voiceless coda
Time-normalized
F2 difference (Hz)
Timepoint

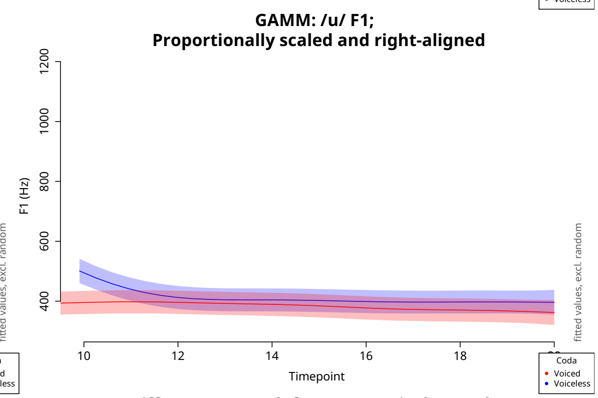

Difference smooth for /ʌ/ F2; voiceless coda
Proportionally scaled and left-aligned
F2 difference (Hz)
Timepoint

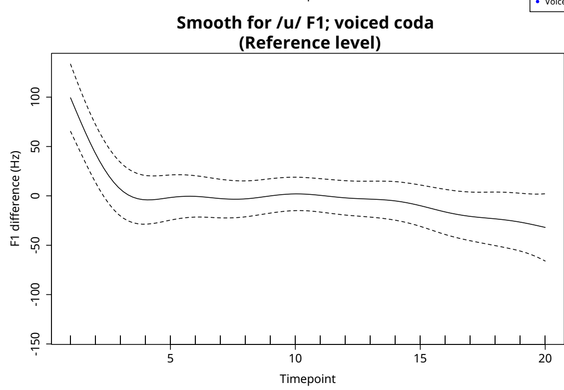

Difference smooth for /ʌ/ F2; voiceless coda
Proportionally scaled and right-aligned
F2 difference (Hz)
Timepoint

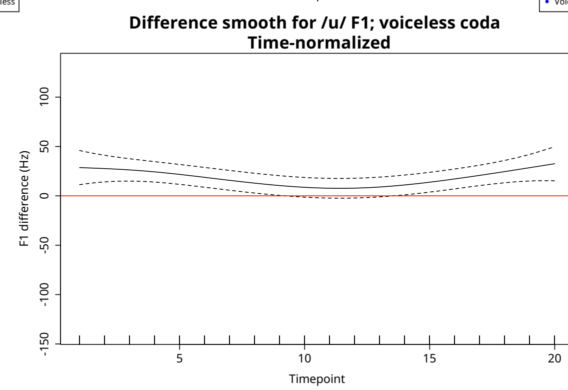
u F1

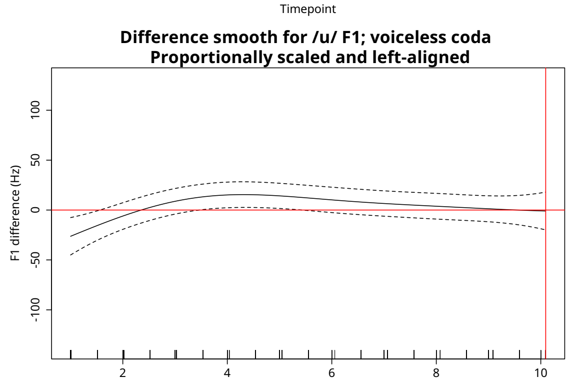

GAMM: /u/ F1;
Time-normalized
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless

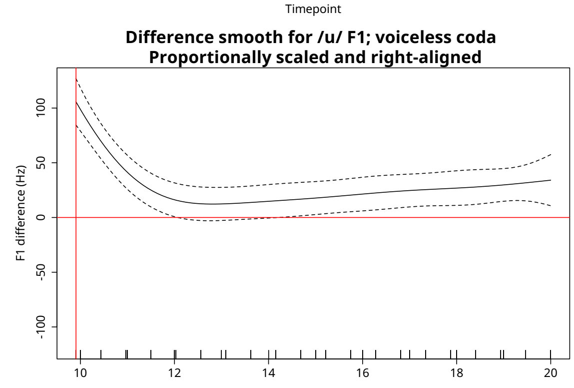

GAMM: /u/ F1;
Proportionally scaled and left-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


GAMM: /u/ F1;
Proportionally scaled and right-aligned
F1 (Hz)
Timepoint
fitted values, excl. random
Coda
Voiced
Voiceless


Smooth for /u/ F1; voiced coda
(Reference level)
F1 difference (Hz)
Timepoint

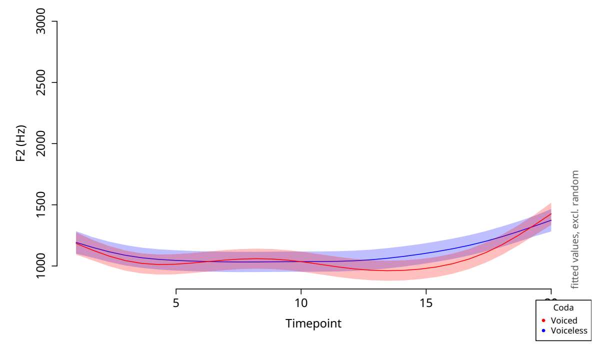

Difference smooth for /u/ F1; voiceless coda
Time-normalized
F1 difference (Hz)
Timepoint


Difference smooth for /u/ F1; voiceless coda
Proportionally scaled and left-aligned
F1 difference (Hz)
Timepoint

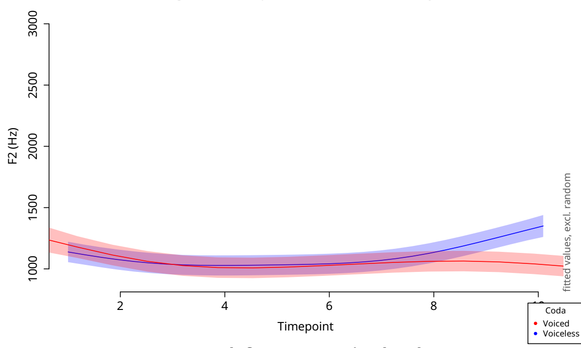

Difference smooth for /u/ F1; voiceless coda
Proportionally scaled and right-aligned
F1 difference (Hz)
Timepoint

u F2

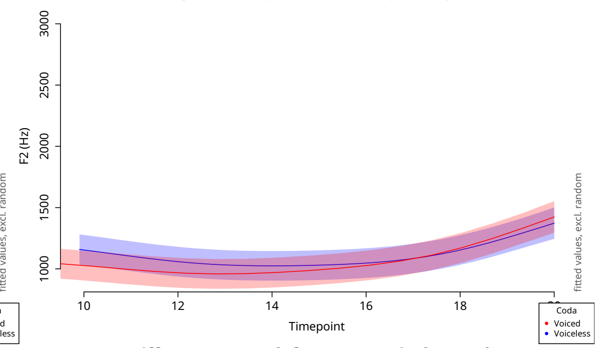
GAMM: /u/ F2;
Time-normalized



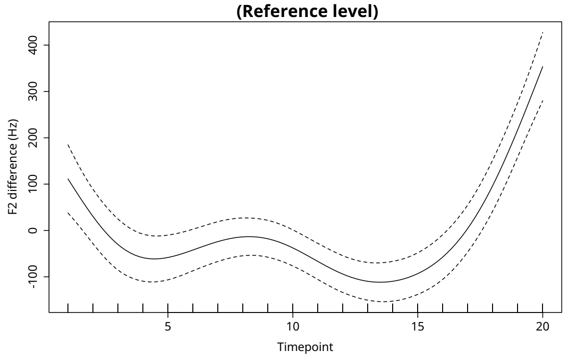
GAMM: /u/ F2;
Proportionally scaled and left-aligned



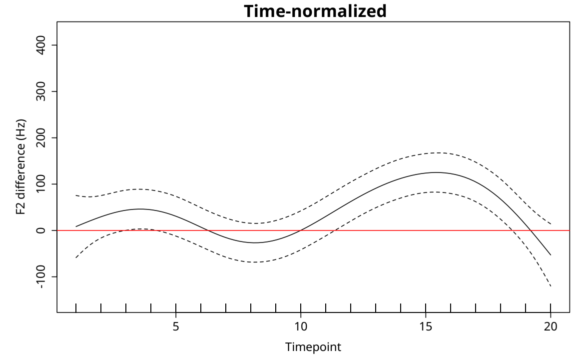
GAMM: /u/ F2;
Proportionally scaled and right-aligned



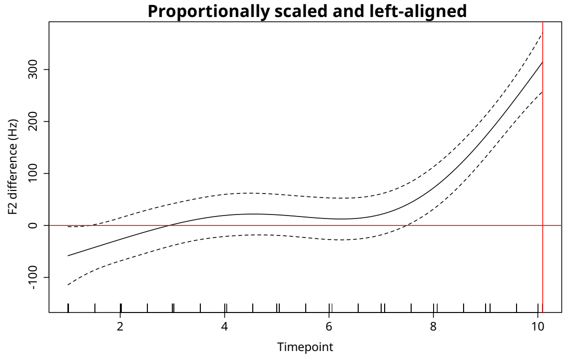
Smooth for /u/ F2; voiced coda
(Reference level)



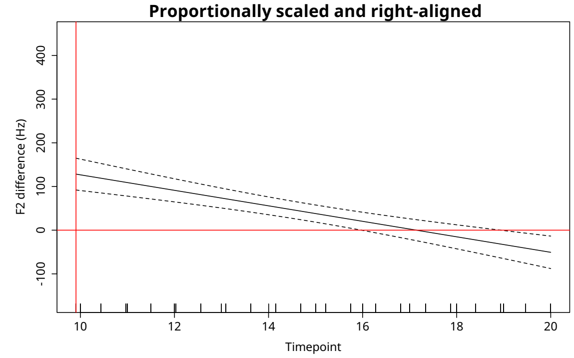
Difference smooth for /u/ F2; voiceless coda
Time-normalized



Difference smooth for /u/ F2; voiceless coda
Proportionally scaled and left-aligned



Difference smooth for /u/ F2; voiceless coda
Proportionally scaled and right-aligned



Appendix H R Scripts

This appendix includes several of the R scripts which I wrote for a number of purposes, primarily to create the various charts and plots throughout this dissertation. I only include here scripts which were wholly my own creation. I cannot ensure the functionality of any of these scripts, or even that they won't break your computer! Use at your own risk. With that being said, I've tried to verbosely annotate the scripts, and tested them afterwards to ensure that they do in fact work as presented – with my data, on my machine.

All of the scripts assume one of two types of `.csv` primary data files. The first contains one token per line, minimally including columns labelled `Vowel`, `Duration`, and `CodaVoice`. The assumed encoding for `CodaVoice` is `0` = voiceless, `1` = voiced, `2` = open syllable. The second `.csv` format, used for timepoint-based calculations and plotting involving formants, should have 20 lines per token with each line contain data from a single timepoint, under the following headings: `Vowel`, `Word`, `Token` (unique identifier), `Timepoint`, `F1`, `F2`, and `F3`. Both files can be automatically generated from properly annotated audio files by using the Praat script `FormantPro.praat` (Xu, 2015), though some tweaking and adjustments may be required. Some formatting options such as font sizes and axis limits were removed from the scripts for readability, as these would likely need to be adjusted on a per-use-case basis dependent on the nature of your dataset and desired output.

The script used to generate the SSANOVAs, as in Appendices B–D, is not included here as it was simply a modification of Wassink (2013), which can be retrieved online from the original source.

R Script: Vowel charts

The following script produces vowel charts as in e.g. Figure 4.2:

```
# Import data
MainData <- read.table (file = "FILENAME.csv", sep=",", header=TRUE)

# Loop through each vowel
Vowels <- levels (Data$Vowel)
for (v in 1:length (Vowels)) {

# Base plot (empty); it is useful to add axis limits when producing
      charts for multiple dialects so that they are identical in scale
plot (0, 0, xlab = "F2 (Hz) (log scale)", ylab = "F1 (Hz)", type = "n",
      log = "x")

# Loop through every vowel
for (v in 1:length (Vowels) ) {

      # Subset data for current vowel
      Data = subset (MainData, Vowel == Vowels[v])
      # Plot mean F2xF1
      Text (mean (Data$mean_F2), mean (Data$mean_F1), labels =
      Vowels[v]) } }
```

R Script: Vowel duration distributions

The following script produced the beanplots (Kampstra, 2008) in e.g. Figure 4.5:

```
# Load beanplot package, import data
library (beanplot)
Data <- read.table (file = "FILENAME.csv", sep=",", header=TRUE)

# Loop through each vowel
Vowels <- levels (Data$Vowel)
for (v in 1:length (Vowels)) {

# Subset by coda
ThisVowel <- subset (Data, Data$Vowel == Vowels[v])
Voiceless <- ThisVowel$Duration [ThisVowel$CodaVoice == 0]
Voiced <- ThisVowel$Duration [ThisVowel$CodaVoice == 1]
Open <- ThisVowel$Duration [ThisVowel$CodaVoice == 2]

# Set up multi-plot frame for 12 vowels in 4 rows, 3 columns; adjust
      for your data
par (mfrow = c (4, 3))

# Make plots, loop back through every vowel
beanplot (Open, Voiced, Voiceless, col = c ("#CAB2D6", "#33A02C",
      "#B2DF8A"), border = "#CAB2D6", names = c ("Open", "Voiced",
      "Voiceless"), ylab = "Duration (ms)", main = Vowels[v]) }
```

R Script: Duration ratios by coda voicing

This script was used to produce the duration ratio plots in e.g. Figure 4.11:

```
# Import data, set up data.frame
Data <- read.csv ("FILENAME.csv")
Data <- Data [order (Data$Duration),]
Vowels <- levels (Data$Vowel)
Durations <- data.frame (matrix (ncol = 4, nrow = length (Vowels)))
colnames (VowelVoiceDur) <- c ("Vowel", "Voiceless Dur", "Voiced Dur",
      "Ratio")

# Calculate duration values for each vowel
for (v in 1:length (Vowels)) {
ThisVowel <- subset (Data, Data$Vowel == Vowels [v])
Durations [v,1] <- Vowels [v]
Durations [v,2] <- mean (ThisVowel$Duration[ThisVowel$CodaVoice==0])
Durations [v,3] <- mean (ThisVowel$Duration[ThisVowel$CodaVoice==1])
Durations [v,4] <- Durations [v,2] / Durations [v,3] }

# Overall duration ratio
Ratio <- mean (VowelVoiceDur[,4])

# Create plot
plot (Durations [,3], Durations [,2], type="n", xlab = "Mean duration
      (ms) before voiced coda", ylab = "Mean duration (ms) before
      voiceless coda", main = "Mean vowel duration, voiced vs.
      voiceless coda")
text (Durations [,3], Durations [,2], labels = Durations [,1])
abline (0, Ratio, col = "red")
legend ("bottomright", paste ("Mean duration ratio:", round(Ratio,3)))
```

R Script: Diphthong trajectories

This script was used to produce the diphthong trajectory plots in e.g. Figure 4.16:

```
# Import data
Data <= read.table (file = "DataFILENAME.csv", sep=",", header = TRUE)
FormantData <- read.table (file = "TimepointDataFILENAME.csv", sep=",",
      header = TRUE)

Diphthongs <- c ("aj", "aw", "ɔj") # diphthongs to be plotted
Vowels <- c ("æ", "ɑ", "ɛ", "e", "ɪ", "o", "u", "ʌ", "ʊ")
DipCols <- c ("blue", "red", "darkgreen") # colours for diphthongs
DipSolid <- c (19, 17, 15) # filled shapes (vowel offset)
DipOpen <- c (21, 24, 22) # open shapes (see Points below)
Codas <- c (1, 0) # coda voice categories
CodaSize <- c (1.5, 1) # size of different-coda shapes
CodaLine <- c (1, 2) # different-coda line types
CodaWidth <- c (2, 1.5) # different-coda line widths
Points <- c (4, 10, 18) # specific timepoints to plot

plot (0, 0, xlab = "F2 (Hz) (log scale)", ylab = "F1 (Hz)", xlim = c
      (2600, 900), ylim = c (1000, 380), type="n", log="x")
for (d in 1:length(Diphthongs)) { # loop through diphthongs
  for (c in 1:length(Codas)) { # loop through coda voicings
    Formants = subset (FormantData, Vowel == Diphthongs[d] & CodaVoice
      == Codas[c])
    F1 <- mean (Formants$F1[Formants$Timepoint == 1]) # F1 at t1
    F2 <- mean (Formants$F2[Formants$Timepoint == 1]) # F2 at t1
    for (t in 2:20) { # loop through all subsequent timepoints
      F1 <- c (F1, mean (Formants$F1[Formants$Timepoint == t]))
      F2 <- c (F2, mean (Formants$F2[Formants$Timepoint == t])) }
    lines (F2, F1, col = DipCols[d], lwd = CodaWidth[c], lty =
      CodaLine[c])
    for (p in 1:length(Points)) { # loop through points to be plotted
      points (F2[Points[p]], F1[Points[p]], pch = DipOpen[d] , col =
      DipCols[d], cex = CodaSize[c])}
    points (F2[20], F1[20], pch = DipSolid[d], col = DipCols[d], cex =
      CodaSize[c]) } }

legend ("bottomright", legend = c (Diphthongs, "Voiced", "Voiceless",
      "20, 50, 90%", "100% dur."), pch = c (DipSolid, 22, 22, 22, 15),
      pt.cex = c(1, 1, 1, 1, 0.7, 1, 1), lty = c (NA, NA, NA, CodaLine,
      NA, NA), col = c (DipCols, "black", "black", "black", "black"))

for (v in 1:length(Vowels)) { # loop through all vowels
  F1 <- mean (Data$mean_F1[Data$Vowel == Vowels[v]])
  F2 <- mean (Data$mean_F2[Data$Vowel == Vowels[v]])
  text (F2, F1, labels = Vowels[v], cex = 1.3) }
```

R Script: GAMMs comparison

This script produces GAMMs difference smooths and smooth comparison plots for each formant per vowel, as in Appendices E–G:

```
# Load required packages and data
library(itsadug) # provides GAMMs functions
library("plotfunctions") # provides <legend_margin> function
Data = read.table (file = "DataFILENAME.csv", sep=",", header = TRUE)
FormantData = read.table (file = "TimepointDataFILENAME.csv", sep=",",
      header = TRUE)

# Make sure that Coda Voice (0,1,2) is interpreted as non-numeric
FormantData$CodaVoice = factor (FormantData$CodaVoice)

# Set up global variables
Vowels = levels(Data$Vowel)
FormantCount = 2 # Add/reduce as desired
Formants=c("F1","F2","F3")

# Loop through every vowel
for (v in 1: length(Vowels)) {

  # Formant timepoint and global means data for current Vowel
  VowelForm = subset (FormantData, FormantData$Vowel == Vowels[v])
  VowelMeans = subset (Data, Data$Vowel == Vowels[v])

  # Set up data frames of formant means at each timepoint in pre-voiced
      and pre-voiceless codas
  VoicedTime = data.frame (matrix (ncol = FormantCount, nrow=20))
  Colnames (VoicedTime) = Formants[1:FormantCount]
  VoicelessTime = data.frame (matrix (ncol = FormantCount, nrow=20))
  Colnames (VoicelessTime) = Formants[1:FormantCount]

  #This part is a bit messy; you need to set the number ("13" in my
      data) to the column in your data just to the left of where F1 is,
      with F2, F3 etc. proceeding to the right
  for (t in 1:20) {
    for (g in 1:FormantCount) {
      VoicedTime[t,g] = mean (VowelForm[,(13+g)][VowelForm$CodaVoice ==
      1&VowelForm$Timepoint == t])
      VoicelessTime[t,g] = mean (VowelForm[,(13+g)][VowelForm$CodaVoice
      == 0&VowelForm$Timepoint == t]) } }

  # Set up variables for scaling and aligning voiceless timepoints to
      Right or Left
  MeansRatio = mean (VowelMeans$duration[VowelMeans$Coda == "t"])/ mean
     (VowelMeans$duration[VowelMeans$Coda == "d"])

  RightOffset = 20-20*MeansRatio
```

```
LeftOffset = 20*MeansRatio
RightRange = 20-RightOffset
LeftRange = 20-LeftOffset

# Loop through formants
for (f in 1:FormantCount) {

  # Load formant timepoint data for current Vowel
  FormNow= as.name(Formants[f])

  # Remove open syllables (coded as CodaVoice = "2")
  VowelForm = subset(VowelForm,VowelForm$CodaVoice!=2)

  # Voiced coda is reference level (voiced CodaVoice = "1" vs.
    voiceless CodaVoice = "0")
  VowelForm$IsVoiceless <- with(VowelForm, ifelse (CodaVoice == "0",
    1, 0))

  # Scale timepoints for Right and Left alignments
  RightForm = VowelForm
  RightForm$Timepoint[RightForm$CodaVoice == 0] = RightForm$Timepoint
    * RightRange / 19 + RightOffset – RightRange / 19
  LeftForm = VowelForm
  LeftForm$Timepoint[LeftForm$CodaVoice == 0] = VowelForm$Timepoint *
    MeansRatio

  # Set up GAMM models: non-scaled, scaled right-aligned, scaled
    left-aligned
  TimeNormModel = bam (RightForm[,13+f] ~ IsVoiceless + s (Timepoint)
    + s (Timepoint, by = IsVoiceless) + s (Timepoint, Speaker, bs =
    "fs", m = 1), data = VowelForm)
  RightModel = bam (RightForm[,13+f] ~ IsVoiceless + s (Timepoint) +
    s (Timepoint, by = IsVoiceless) + s (Timepoint, Speaker, bs =
    "fs", m = 1), data = RightForm)
  LeftModel = bam (LeftForm[,13+f] ~ IsVoiceless + s (Timepoint) + s
    (Timepoint, by = IsVoiceless) + s (Timepoint, Speaker, bs = "fs",
    m = 1), data = LeftForm)

  # 7 combined plots for each formant/vowel
  par(mfrow=c(4,2))
  plot (c (0, 1), c(0, 1), ann = F, bty = 'n', type = 'n', xaxt =
    'n', yaxt = 'n')
  text (x = 0.5, y = 0.5, paste (Vowel[v],Formants[f]), cex = 7, col
    = "black")
  plot_smooth (TimeNormModel, view = "Timepoint", cond = list
    (IsVoiceless = 1), rug = F, rm.ranef = T, col = "blue", main =
    paste0 ("GAMM: /",Vowel[v],"/ ",FormNow,"; \n Time-normalized"),
    ylab = paste (FormNow,"(Hz)"))
  plot_smooth (TimeNormModel, view = "Timepoint", cond = list
    (IsVoiceless = 0), rug = F, rm.ranef = T, col = "red", add = T)
  legend_margin ("bottomright", title = "Coda", legend = c ("Voiced",
    "Voiceless"), col = c ("red", "blue"), pch = 20)
  plot_smooth (LeftModel, view = "Timepoint", cond = list
    (IsVoiceless = 1), rug = F, rm.ranef = T, col = "blue", main =
```

```
  paste0 ("GAMM: /",VowelIPA[v],"/ ",FormNow,"; \n Proportionally
  rescaled and left-aligned"), ylab = paste (FormNow, "(Hz)"), xlim
  = c (1, LeftOffset))
plot_smooth (LeftModel, view = "Timepoint", cond = list
  (IsVoiceless = 0), rug = F, rm.ranef = T, col = "red", add = T)
legend_margin ("bottomright", title = "Coda", legend = c ("Voiced",
  "Voiceless"), col = c ("red", "blue"), pch = 20)
plot_smooth (RightModel, view = "Timepoint", cond = list
  (IsVoiceless = 1), rug = F, rm.ranef = T, col = "blue", main =
  paste0 ("GAMM: /",VowelIPA[v],"/ ",FormNow,"; \n Proportionally
  scaled and right-aligned"), ylab = paste (FormNow,"(Hz)"), xlim =
  c (RightOffset, 20))
plot_smooth (RightModel, view = "Timepoint", cond = list
  (IsVoiceless = 0), rug = F, rm.ranef = T, col = "red", add = T)
legend_margin ("bottomright", title = "Coda", legend = c ("Voiced",
  "Voiceless"), col = c ("red", "blue"), pch = 20)
plot (TimeNormModel, select = 1, main = paste0 ("Smooth for
  /",VowelIPA[v],"/ ",FormNow,"; voiced coda \n (Reference
  level)"), ylab = FormantLabel)
plot (TimeNormModel, select = 2, main = paste0("Difference smooth
  for /",VowelIPA[v],"/ ",FormNow,"; voiceless coda \n Time-
  normalized"), ylab = FormantLabel)
abline (0, 0, col = "red", lty = 1)
plot (LeftModel, select = 2, main = paste0 ("Difference smooth for
  /",VowelIPA[v],"/ ",FormNow,"; voiceless coda \n Proportionally
  scaled and left-aligned"), xlim = c (1, LeftOffset), ylab =
  FormantLabel)
abline (h=0, col="red", lty=1)
abline (v=LeftOffset, col = "red", lty=1)
plot (RightModel, select = 2, main = paste0 ("Difference smooth for
  /",VowelIPA[v],"/ ",FormNow,"; voiceless coda \n Proportionally
  scaled and right-aligned"), xlim = c (RightOffset, 20), ylab =
  FormantLabel)
abline (h = 0, col = "red", lty = 1)
abline (v = RightOffset, col = "red", lty = 1) } }
```