Multidimensional stylistic variability of diphthongization in Québec French: a corpus study

Author: Liam Bassford
Supervisors: Peter Milne, Morgan Sonderegger

A thesis presented in support of the degree of Bachelor of Arts (Honours) in Linguistics

McGill University
Department of Linguistics
April 21, 2015

0. Abstract

Phonetic style-shifting can be seen, under various names, in many linguistic studies. Two commonly-studied dimensions are "attention to speech" and "clear speech" or "hyper/hypospeech." The former is common in sociolinguistics papers; the latter is more often used by acoustic phoneticians. Intuitively, and by many acoustic measures, these dimensions line up well with each other; for example, clearer speech and more attentive speech are both associated with lower speaking rates and less segment reduction. Indeed, socially stigmatized variants often require less articulatory effort (see Kroch 1978). However, there is a way to put the continua into conflict: by studying the behavior of stigmatized variants that require more articulatory effort than the standard. One such variable is the diphthongization of long vowels in Québec French; the diphthongized variants require greater energy expenditure, due to increased articulator movement when compared to the monophthongized standard. By studying the behavior of the long /ɛ:/ in Québec French, using corpus data collected by Milne (2013), we find that the variable, paradoxically, becomes suppressed in attentive speech and expressed more often in clear speech, suggesting that the two continua are not identical. Furthermore, gender differences in the corpus show a potential way to reconcile the apparent conflict: compared to men, women in the corpus realize the diphthongized (nonstandard) variant less frequently by discrete perceptual measures but more radically by a gradient acoustic measure.

1. Introduction

1.1 Stylistic variability in speech

Individual variability in speech has long been a vital object of linguistic study. Ronald Wardhaugh (2006) summarizes some of the basic observations about individual variability in speech:

> "When we look closely at any language, we will discover time and time again that there is considerable internal variation and that speakers make constant use of the many different possibilities offered to them. No one speaks the same way all the time and people constantly exploit the nuances of the languages they speak for a wide variety of purposes." (p. 5)

In Wardhaugh's view, there is a temptation in many traditions of linguistics towards simplifying models of linguistic phenomena. This tendency is convenient when investigators are attempting to formulate basic rules about how a language operates. However, certain observations cannot be safely ignored--namely, the fact that no formulated rule about any aspect of language is completely regular; the closer one investigates the application of the rule, the more likely one is to find "wrinkles" and irregularities in any given formula (pp. 4-5).

The obvious question to Wardhaugh is: *To which ends do speakers exploit these nuances and uncertainties in their language?* For many years, the consensus was that speakers were *not* exploiting this natural variability. Although variation was recognized, linguists were puzzled as to the source and distribution of this variation. In particular, the techniques for measuring variation in speech were so unrefined that nobody could draw useful conclusions. Labov (1972, ch. 2) contains an early attempt at characterizing speech variability, the "department-store study" of 1962. His goal was to determine whether there was social stratification of the well-known (r)

variable in New York City English. The variable is the presence or absence of [r] in the coda position of a syllable following a vowel. On the subject of the (r) variable, Labov quotes a previous explanation for its distribution--free variation: "The speaker heard both types of pronunciation about him all the time, both seem almost equally natural to him, and it is a matter of pure chance which one comes to his lips" (Hubbell 1950, p. 48). Labov had become convinced that the variable was distributed in a manner that reflected the pervasive social stratification in New York City. He did, in fact, find that this stratification was mimicked in the realization of the (r) variable--not just in broad class distinctions like that between a doctor and a janitor, but also in fine-grained distinctions, like slightly different-ranking employees of the same department store.

Style-shifting refers to an axiom in the sociolinguistic tradition that individual speakers do not control just one style of speech, but rather a range of styles which they can modulate based on varying circumstances (Wardhaugh 2002, 19). Bell (1984) sees style shifting as a function of *audience design*—it is based on a speaker's accommodations to their addressee, and to a lesser extent accommodations to third-party "referees," especially in mass communication. Wardhaugh (2002) distinguishes style-shifting on the one hand from dialect differences (which characterize *interspeaker* variation), and on the other hand from *register* (features associated with "discrete occupational or social groups"—for example stockbrokers, mountain climbers, or sports fans) (pp. 51-52).

1.2 Attention to speech: a sociolinguistic measure

The most commonly-studied dimension of style is a measure of the speaker's *attention to speech* (Labov 1972), and will, for the rest of the paper, be referred to as "attentive speech," regardless of the terminology used in sources. The specifics of this dimension are outlined in great detail in Labov's "The Isolation of Contextual Styles" (1972, ch. 3). Labov characterizes a number of contextual styles which influence the realization of stratified sociolinguistic variables, which stratification he demonstrated in "The Social Stratification of (r) in New York City Department Stores" (1972, ch. 2). The first distinction he makes is between *careful speech*, the speech of a formal sociolinguistic interview; and *casual speech*, which is speech realized outside an interview--for example, on the street, on the subway, or in bars. Casual speech, in this model, cannot be captured by an overt interview due to the effects of observation. However, he outlines a number of methods to capture casual speech; some of these are:

1. Incidental speech outside the context of the interview, such as the subject offering a cup of coffee or beer to the interviewer
2. Speech to a third party, such as an interrupting telephone call or a curious child
3. Speech which digresses significantly from the initial interview question

Labov identifies several "channel cues" to separate casual speech from careful speech. Casual speech is associated with changes in tempo, volume, breathing, and pitch range; Labov does not specify a direction for any of these changes, but notes that laughter (associated with breathing changes) increases in casual speech.

Labov cites Mahl (1972), a study with four conditions. Subjects were given headphones with levels of white noise loud enough that they were no longer able to hear themselves, or told to face away from the interviewer, or both, or neither. He found that the realization of the English /ð/ phoneme as a stop rather than as a fricative was most common when subjects were given the white noise *and* told to face away from the interviewer. After that, stopping was realized less frequently (in decreasing order) when facing the interviewer with noise, facing away without noise, and facing the interviewer without noise. Labov uses this all as evidence to hypothesize that speech style varies along a single dimension, that of attention to speech. For the purposes of the study to follow, we will be focusing on the attention-to-speech dimension, as it is acknowledged to influence style-shifting, commonly used in sociolinguistics, and conducive for use in our particular data set.

That is not to say that the attention-to-speech continuum is universally accepted as the only means of style-shifting; criticisms of the measure have been raised, and other viable alternatives have been proposed. The main direct challenge to Labov's theories of variability in speech has been the *audience design* theory proposed by Bell (1984). Revisiting the data of the Mahl (1972) study, Bell notes that although the noise *did* seem to influence the stopping of dental fricatives, it had a weaker effect than that of facing away from the interviewer, suggesting that the more important factor is the subject's awareness of the interviewer. Labov's taxonomy of speech styles is useful to Bell only as a methodological construct, not a theoretical one; many of his scenarios are not reflective of any speech likely to be seen outside of the context of a sociolinguistic interview, such as the reading of minimal pairs (p. 150). Bell points out that stylistic variation within a speaker is nearly always a *subset* of the total social (interspeaker) variation within a speech community, except in cases of hypercorrection and sometimes in cultures with a large degree of ritualized interpersonal deference (e.g. in the case of Tehrani Persian) (151-155). Therefore, in his view, any view of intraspeaker variation must be defined in the same terms as intraspeaker variation. The paradigm of audience design starts with interspeaker variation at the core, and the intraspeaker variation modified by the speaker in response to the following interlocutors, in order of importance:

1. Addressee—the listener being directly addressed by the speaker
2. Auditor—a listener whose presence is known and ratified, but who is not addressed
3. Overhearer—who is known but neither ratified nor addressed

Eavesdroppers—those who are not known, ratified, or addressed—do not influence the speaker's style-shifting. He cites numerous studies to support the finding that subjects tend towards a style based on their audience, by comparing their speech in a peer group to that in a formal interview with an outside investigator (Bickerton 1980, Douglas-Cowie 1978, Russell 1982) (pp. 163-4). In Coupland (1984), a travel agent's speech was found to converge or even hyperconverge with her clients' speech in multiple variables including intervocalic voicing of (t), based on the clients' class. He revisits Labov's department store study, finding that the largest differences in the (r) variable were *between* department stores, despite the fact that the employees at Saks were not of a higher social class than their counterparts at Macy's or S. Klein, because the social classes of

each store's *clientele* were different, and his subjects were accommodating the addressee (i.e. Labov) differently based on his expected social class.

Zwicky (1972) is another early investigation into the nature of "casual speech," which, as with Labov, he contrasts with "careful speech." According to Zwicky, casual speech is often, but not necessarily, produced at faster speaking rates—in line with Labov's (1972, ch. 3) prior assessment of speaking rate as a *channel cue* for attention to speech. Speaking rate cannot be the sole indicator of formality--it is possible to give the impression of formality at rapid rates or casualness at slower rates. Rather, he enumerates many other casual speech processes. In his view, these processes are often phonetically natural. It is rare that processes of formal speech become physically impossible to articulate, but they usually become more difficult to articulate as speaking rates increase. They fall into two categories: processes which increase facility (e.g. assimilation, neutralization) and brevity (e.g. degemination, monophthongization).

1.3 Clear speech: an acoustic measure

Clear speech is a dimension of speech variation often studied by phoneticians, and it has different theoretical foundations from the style-shifting measures described in Labov (1972). It is an acoustic dimension of speech variability with two poles: *citation speech*, an unmarked form of speech; and *clear speech*, which is marked. Lindblom (1990) provides a comprehensive theoretical foundation of the clear-speech dimension, though he refers to it as the *H&H theory*. "H&H" refers to *hyperspeech* and *hypospeech*, which are also described as *listener-oriented* and *speaker-oriented* speech, respectively. The initial issue being explored is the *invariance problem*, namely, that it is difficult or impossible to define a linguistic category consistently without knowledge of its context, given the enormous range of intra-speaker variation in any given category. To Lindblom, speech perception is a process of discrimination between lexical items. The speaker's usage of their articulatory functions is aimed towards the goal of facilitating this discrimination for the listener ("hyperspeech"). This goal is in conflict with another goal, that of conserving effort, so when constraints on articulation lessen, the articulation tends towards more economic motion ("hypospeech"). The constant process of balancing these goals creates another dimension of speech variability, between hyperspeech and hypospeech, in the same way that a speaker's level of attention to speech creates a continuum between formal and vernacular speech.

The H&H theory is also somewhat reflective of Bell's audience design theory, in that it asserts that speakers have a choice to accommodate the listener. However, the conflict in H&H is not a purely social distinction between sociolinguistically stratified ways of speaking, but rather between the more selfish biological tendency to accommodate energy-conserving articulatory behavior and the more pragmatic social tendency to aid the listener in comprehension of speech.

Another important investigation into the clear speech continuum was Moon and Lindblom (1989). In this study, the investigators elicited front vowels in two CVC contexts: /h_d/, to sample the vowel space as a control, and /w_l/ as the test context. Because [w] and [ɬ] (the coda realization of /l/) both have a velar place of articulation, and the test vowels were front vowels, this resulted in very wide V-to-C transitions in articulatory space. Vowel durations were controlled by the use of different-length words, e.g.: *wheel, will, well, wail; wheeling, willing;*

Wheelingham, Waillingby. The longer words would have relatively shorter vowel durations. Subjects were told to recite a word list in a natural manner first (citation form), and then given the direction to speak as if they were speaking to someone who was not a native English speaker. Given the wide articulatory movements necessary to go from back-to-front-to-back, it was hypothesized (and then confirmed) that subjects would show *undershoot* of their front vowels, that (for example) the F2 value of [i] in *wheel* /wil/ would be lower than that in the control form, *heed* /hid/, where the transitions are less drastic. Furthermore, in the multisyllabic words, there would be further undershoot owing to the even shorter vowel duration available to fully articulate the [i]. Finally, the clear speech condition was shown to mitigate the effects of undershoot, with clearly-enunciated words having formant values closer to the control /h_d/ context than the citation-form /w_l/ context. Lax vowels /ɪ ɛ/ were more prone to undershoot than tense vowels /i eɪ/. Moon and Lindblom (1989) is evidence for the existence of a clear speech continuum of variation, and specifically identifies the absence of undershoot as one of the features of that continuum—in clear speech, vowels are more likely to be realized with their full value, rather than a value influenced by the position (in this case, backness) of the surrounding segments.

As one application, the H&H model fits neatly with Zwicky's observation that neutralization of phonemic distinctions is a common process in casual speech: As the necessity of making finer phonemic distinctions becomes less necessary (for whatever reason), these distinctions may begin to disappear, if their disappearance would lessen articulatory effort. To apply this to a sociolinguistic situation, one English variable, which we will revisit frequently, would be English dental stopping--Labov's (th) and (dh) variables, where /θ/ and /ð/ are realized as [t] and [d], respectively (Labov 1972, ch. 3). If a given speaker's idiolect is prone to this process, by Lindblom's reckoning, it would be more likely to show up where the distinction between /θ/ and /t/ is unimportant.

Smiljanic and Bradlow (2009) investigated the notion of clear speech in a review of prior studies. They aimed to determine two things: which phonetic correlates are representative of clear speech, and which features actually contribute to greater intelligibility. They first define clear speech roughly as speech that is slower, louder, and more "exaggerated." Like infant-oriented or computer-oriented speech, clear speech is goal-oriented; it sacrifices economy of effort for intelligibility to the listener—in other words, it deals with the same conflicting goals as Lindblom's H&H theory, measuring speaker-oriented or listener-oriented speech. It is generally effective at increasing intelligibility, both for native and non-native listeners, or normal-hearing and hearing-impaired listeners. However, clear speech is not necessarily intelligible speech, and vice versa. The use of clear speech cannot overcome the total loss of ability to distinguish pairs of sounds; for example, if a listener has sloping hearing loss (which makes the distinction between sibilants imperceptible), clear speech does not aid in the perception of sibilants. It is important to note that clear speech is typically defined based on what speakers *actually do* in an attempt to sound clearer, not which factors truly help listeners to comprehend.

Clear speech is correlated with a lower speaking rate, wider pitch range, higher volume (SPL), and more salient stop releases. At least in English, expanded vowel spaces are also seen, as a consequence of either decreased undershoot or a retargeting of gestures to focus on vowels. In fact, some talkers are naturally more intelligible at a baseline level than others, and they tend to have larger vowel spaces on average. Slower speaking rates entail not only longer segments, but also more frequent and longer pauses; however, despite being ubiquitous in clear speech, speaking rate alone does not consistently benefit intelligibility. Clear speech strategies have some variation crosslinguistically. Compared to English, Croatian has long vowels that vary only in duration (not duration *and* quality); it also has prevoiced and short-lag stops, where English has short-lag and long-lag stops. In Croatian clear speech, it was found that the duration distinction between long and short vowels was emphasized more than in English clear speech, and that where English extended the lag of its long-lag ("voiceless") stops, Croatian increased the length of prevoicing in its voiced stops (pp. 5-11).

Gahl, Yao, and Johnson (2011) found that a speaker's choice to hypo-articulate or not can be determined by many other factors, including a word's frequency and neighborhood density. Frequency and neighborhood density were found to affect word duration--a correlate of speech clarity--significantly; neighborhood density significantly affected the other studied correlate, vowel dispersion, as well. In the study, more frequent and high-density words tended to be realized with shorter durations and less dispersion. However, they note that neighborhood density has normally been found to have the opposite effect, with words in more dense environments becoming more clearly articulated; they attribute the difference to methodological differences from previous studies (pp. 10-15).

1.4 Reconciling the two continua

Kroch's "Toward a Theory of Social Dialect Variation" (1978) is a point of contact between the two paradigms. Although he does not examine the clear-speech literature specifically, he is one of the earlier sociolinguists to focus on the articulatory effort aspects of stratified speech variables. Kroch's thesis in this paper is that "popular dialects exhibit their greater susceptibility to phonetic conditioning in such features as simplified articulation, replacement or loss of perceptually weak segments, and a greater tendency to undergo 'natural' vowel shifts" (19)--in other words, vernacular or stigmatized variants of a stratified variable are often articulatorily easier to pronounce. This thesis echoes the aforementioned observation from Zwicky (1972) that casual speech processes tend to be phonetically natural. His evidence includes a number of phenomena from across several languages, including Spanish debuccalization[1], English dental stopping,[2] and consonant cluster reduction in Québec French. Many of these phenomena are seen to pattern similarly in other studies. For example, English dental stopping figures prominently in Labov (1972, chapter 3) and Mahl (1972), and indeed is found more often in less attentive speech. However, as we will see, his theories are not applicable to all stratified variables.

[1] /s/ → [h] or ø / _σ
[2] /θ ð/ → [t d]

We can see from the literature that attention to speech and clear speech, though studied in separate bodies of literature, have many similarities and appear to be somewhat correlated, with attentive speech corresponding to clear speech and casual speech corresponding to citation speech. Furthermore, many processes seen in casual speech are natural and phonetically simplifying, and thus more likely to be avoided in clear speech. It is therefore tempting to conclude that clear-speech literature and attentive-speech literature are attempting to quantify the same phenomenon, a single continuum. However, there are variables which suggest that clear speech and attentive speech are not entirely identical--if a stigmatized or vernacular variant takes more articulatory effort than the standard variant, it should be suppressed in the attentive-speech model but exaggerated in the clear-speech model.

1.5 In this thesis

Several potential candidates present themselves for a variable to counter Kroch. Labov's work on New York City English (1972, ch. 3) described the (eh) and (oh) variables, in which the vowels /æ/ and /ɔ/ are split into marginally-phonemic tense/lax pairs.[3] They are similar in character, except that (eh) is for front vowels and (oh) is for back vowels. In both cases, the tense variant is an ingliding diphthong, with [ə] as the second element. The first element is a front or back vowel, realized on a spectrum of height from [ɪə] and [ʊə] at the extreme high end to [æ(ə)] and [ɑ(ə)] at the low end; realizations with a higher onset are considered more vernacular. If we are to follow Zwicky's observation that monophthongs tend to replace diphthongs in casual speech processes, then the characterization of (eh) and (oh) runs contrary to Kroch's prediction that vernacular variants are necessarily simpler to articulate. What we have in the case of (eh) and (oh), instead, is two variables where, as their realizations become more vernacular, they also become more challenging to articulate, due to the larger physical motion required in reaching the higher vowels.

We will be investigating a similar variable, diphthongization of long vowels in Québec French. The diphthongization variable is similar because, like with (eh) and (oh), the diphthongized variant is the stigmatized one, and, as we will see, the variants which are more extreme in their articulatory motion are more obviously vernacular, while those closer to monophthongs are more standard. After discussing the variable in depth, we will move on to an experiment designed to answer the following questions:

1. Do the clear-speech and attention-to-speech continua differ? If so, how?
2. Is one continuum more accurate than the other in measuring the widespread natural variability in natural speech? That is, if we put them in a situation where they conflict, will one prevail?

[3] For the (eh) variable, a minimal pair would be *draggin'* [dreəgɪn] / *dragon* [drægɪn]. For the (oh) variable, a minimal pair would be *caught* [koət]~[kɔət] / *cot* [kɑt]

2. Data

2.1 The variable

In order to get a good picture of the interactions between these two speech continua, we needed to find a variable which would put them into conflict. Firstly, we have seen that many of the same changes that occur in less attentive styles (such as speaking rate) also occur in less clear styles, and vice versa. Furthermore, stylistic variation and social variation are interrelated, and variables which are more common in the speech of lower-class subjects are also more common in casual speech overall—as Labov (1972) puts it, a "casual salesman" and a "careful pipefitter" may have similar speech patterns (p. 151). We have seen that casual speech processes tend to be phonetically natural—that is, they tend to simplify physically difficult segments (Zwicky 1972). In line with this observation, we have also seen speculation that, in stratified variables, the stigmatized variant is likely to be simpler to articulate (Kroch 1978). Taken all together, it seems quite likely that clear speech and attentive speech are ultimately referring to the same process.

However, this is not a logical necessity; it is possible to imagine a case where the stigmatized variant of a socially-stratified variable requires *more* articulatory effort. If we found such a variable, we would intuitively expect realization of the stigmatized variant to be less frequent in attentive speech (in the sociolinguistic sense), but more frequent in clear speech (in the phonetic sense, because the stigmatized variant is also the more effortful variant). And if those predictions were borne out, then we could demonstrate that the two continua are not the same thing; if they were *not* borne out, and the variable in question behaved similarly in high-attention and high-clarity styles, then we may be able to choose which continuum is more important in determining the distribution of the variable, by determining which continuum the variable patterns with more conventionally. Kroch (1978) does not discuss any such changes; as a possible counterexample to his theory, he cites the history of New York City English /r/-deletion, which is theorized to have originated as an upper-class innovation. However, this is a case of the prestige dialect succumbing to articulatory simplification, not of a vernacular dialect avoiding it.

Such variables are admittedly more difficult to find, but one variable that qualifies is the diphthongization of long vowels in Québec French. This process has been described in numerous sources, sometimes conflicting on basic points (Walker 1984, Santerre & Millo 1978, Ostiguy & Tousignant 2008, Côté 2005). It is generally agreed that certain (not all) Québec French vowels can be realized as short or long. Côté identifies /i e ɛ a y ø u ɔ/ as the eight vowels that can undergo lengthening. Santerre and Millo identify pairs of short and long vowels (p. 173):

Short ~ Long

fait /fɛt/ 'fact' ~ *fête* /fɜt/ 'party'
patte /pat/ 'paw' ~ *pâte* /pɑt/ 'paste'
notre /nɔtʀ/ 'our' ~ *(le) nôtre* /notʀ/ 'ours'
jeune /ʒœn/ 'young' ~ *jeûne* /ʒøn/ 'fast (avoidance of eating)'

Walker posits that /o ø ɑ/ are always intrinsically long, but otherwise agrees with Côté's assessment. Mackenzie and Sankoff mostly follow Walker and Côté, but do not identify /a/ as capable of lengthening, and *do* include /œ/ as a potentially-long vowel.

Despite the disagreement on how to represent the long vowel system of Québec French, the literature generally agrees that lengthening can happen for two reasons--one allophonic, one quasi-phonemic. The vowel could appear before a lengthening consonant (*consonne allongeante*), in a word-final closed syllable. Vowels in a word-final syllable ending in /ʀ z ʒ v/ (*père* 'father'*, chaise* 'chair'*, neige* 'snow'*, grève* 'strike [labor]'), or certain consonant clusters such as /vʀ/ (*chèvre* 'goat'), are subject to the allophonic lengthening process (Mackenzie & Sankoff 2010). There is also a class of words with inherently- or historically-long vowels, where the lengthening is an artifact of an old sound change. This can be compensatory lengthening for a degeminated or deleted consonant, e.g. *baisse* 'decrease (n)' or *même* 'same' (< **mesm*) (Mackenzie & Sankoff); it is often indicated orthographically with a circumflex, <ê> or <â>. Côté (2005) notes that long vowels are usually found in a stressed syllable and never word-finally; Walker (1984), however, gives examples of long vowels in pretonic position, such as *fêter* /fɛ:te/ 'to celebrate.'

Diphthongization can be seen as a form of fortition; it increases distinctiveness of long vowels as opposed to their short counterparts (Walker 1984). Walker describes the process as the addition of an offglide to the monophthong, which can be /j/, /w/, or /ɥ/ depending on the frontness and rounding of the first element. Short vowels (particularly /ə/) and /a:/ never diphthongize, but /ɛ:/, in addition to its expected realization as [ɛj], can sometimes be realized as [aj], with a lowered and slightly backed onset. Mackenzie and Sankoff (2010) additionally noted that the diphthongization of high vowels /i:/, /y:/, and /u:/ is contested between sources, but in their own study, they did find significant changes in vowel quality between the onset and glide of long high vowels.

Despite the largely-allophonic basis for vowel length in Québec French, owing to the presence of historically-long vowels, there are actually a handful of minimal pairs that may indicate a long/short distinction: *fait* /fɛt/ ‘fact’ ~ *fête* /fɛ:t/ ‘party’; *mettre* /mɛtʀ/ ‘to put’ ~ *maître* /mɛ:tʀ/ ‘master’ (Santerre & Millo 1978); *tache* /taʃ/ ‘stain’ ~ *tâche* /tɑ:ʃ/ ‘task’ (Dumas 1974 in Gess 2008). Dumas asserts that these minimal pairs do not have general or productive status within the language; rather, they are inherited from a very baroque system of compensatory lengthening which is no longer relevant in modern French. Mackenzie and Sankoff (2010) note that this categorization is common in the literature. In other words, the contrast between /ɛ/ and /ɛ:/ can be seen as only marginally phonemic; the literature generally makes the simplifying assumption that length is *not* phonemically contrastive in Québec French. We will keep in mind the distinction between historically-long and allophonically-long /ɛ:/, in case they behave differently in practice, but we will not assume them to be different *a priori*.

For the purposes of this study, we will be looking only at the /ɛ:/ vowel. /ɛ:/.There are two reasons for this restriction: first, /ɛ:/, apart from being very common and uncontroversially subject to lengthening, has a more dramatic range of variability than any other long vowel; it is the only one whose *onset* often changes in value from its short counterpart. It also has the benefit of appearing in both allophonic and historically-long contexts, with a handful of minimal pairs. Finally, diphthongization of /ɑ:/, another strong candidate for study, had been found to be in

decline in 1984, unlike /ɛ:/, which showed lowering and backing of the onset (to [a]) from 1984 but no decline in diphthongization (Mackenzie and Sankoff 2010, 96-97). Given the time elapsed between the 1984 recordings and the 2011 recordings, we would expect this decline to continue. Throughout the paper, we may use the terms "allophonic" to refer to tokens of /ɛ:/ before a lengthening consonant, and "phonemic" to refer to the historically-long tokens which occur before something besides a lengthening consonant.

2.2 The ANQ corpus

The data came from a corpus consisting of 60 hours of speech from the Assemblée Nationale du Québec (ANQ) recorded during proceedings of the 39th Legislature in 2011, as prepared by Milne (2013). Milne had already force-aligned the data. Forced alignment is a process by which segments occurring in natural, continuous speech are automatically labelled using probabilistic models based on the acoustic correlates of phonemic segments. Given the time-consuming nature of annotating segment boundaries by hand, forced alignment offers a speedier alternative. The force-aligner starts with a training set of speech which is annotated, segment-by-segment, with start and end times for every segment in the data. It is then given its test set as input—that is, a sound file and its orthographic transcription. Its output is a mapping of the orthographic transcription to a broad (phonemic) IPA transcription, and the start and end times for each segment in the broad transcription (Milne 2013).

The ANQ corpus contains 105 speakers, 38 female—although only 41 male and 20 female speakers ended up having usable tokens of /ɛ:/, for a total of 61 speakers studied. The speech had already been coded for one of two styles; it consists of both impromptu speech and reading of prepared documents. This feature is rare in non-laboratory corpus data, but advantageous for the study because it explicitly indicates the attention-to-speech continuum being studied, approximating the reading-passage and word-list levels of speech found in Labov (1972, ch. 3). The baseline level of formality in the corpus is quite high--it is all realized within the rather ritualized environment of the ANQ, and so while there is no formal linguistic interview setting in the data, the context is not casual; the speakers are certainly aware that what they are saying is being recorded and scrutinized by both peers and outsiders. None of the speech is coded explicitly for "clarity," but the level of speech clarity in the data can be inferred by measuring several speech variables in the words and sentences surrounding the tokens, including pitch variance, speaking rate, and average segment duration of tokens. There are language-specific ways to measure clear speech as well; most obviously, the very same ANQ corpus shows a 72.32% rate of consonant cluster reduction (Milne 2013, 76), so the presence of a greater proportion of non-reduced clusters in a given sound file can indicate clearer speech.

The ANQ corpus has a wide sampling of regional subdialects. Much of the previous investigation into the diphthongization variable has been focused on the Montréal dialect (e.g. Mackenzie and Sankoff 2010, Santerre and Millo 1978). The ANQ, on the other hand, contains representatives from every corner of Québec. Though it would probably be impossible to survey regional dialect differences, given the small amount and uneven distribution of tokens among speakers, the focus is less restrictive than in previous work, and although the results may be less

directly comparable to each other, they can also be generalized to the rest of Québec, not just the Montréal area.

The limitations of the corpus should be noted, however. Female speakers are underrepresented in the data—out of the 61 speakers with relevant /ɛ:/ tokens, only 20 are female. The subject matter is limited and stylized, so the variable being studied is most often found in either very frequent words (*être, même*) or in a limited set of words relating to politics and finances. Certain words such as *financière*, *enquête*, and *budgétaire* occur quite frequently, for example. Finally, although it was done professionally for the purposes of being intelligible to any listener, the corpus was not originally recorded under laboratory conditions with acoustic analysis in mind, so occasionally a token may be confounded because of excessive reverb in the room, or a given speaker's distance from the microphone. This unfortunately negates the usefulness of sound pressure level as a measure of clear speech, and resulted in some tokens being thrown out which would probably have survived had they been recorded in a lab setting.

2.3 Methods

The values of the /ɛ/ tokens capable of undergoing diphthongization were measured in Praat (Boersma & Weenink. Since the original forced aligner did not distinguish between /ɛ/ and /ɛ:/ phonemes, nor annotate allophonically-lengthened /ɛ/, we used a script to pick out only words which had /ɛ/ tokens in the final closed syllable, whose coda was /ʀ z ʒ v/ or a consonant cluster beginning with a lengthening consonant such as /vʀ/. Additionally, it picked out a whitelist of words which were written with <ê> in the orthography--a telltale sign of a historically-long /ɛ:/. See the Appendix 1 for a full whitelist.

The measurements were taken semi-automatically: A script took measurements of many parameters, including F1, F2, and F3, in Praat, for the onset and glide of every token. Measurements of diphthongs can be taken in a number of ways, but one common practice is to measure at two points, towards the beginning and end of the vowel, and equidistant from the center (e.g. Thomas 2010, pp. 151-152). For this study, the onset measurement was taken at a single point 25% of the way into the vowel, and the glide measurement at a point 75% into the vowel. Throughout the annotation process, the author manually moved some segment boundaries from their original force-aligned positions, in the occasional case that they were obviously inappropriately-placed.

Formant values are realized in part by a speaker's intentions to realize a phoneme, combined with their sociolinguistic tendencies to realize that phoneme a certain way. However, raw formant values can be misleading in sociolinguistics, especially when working with large sets of data with multiple speakers. This is because formant values are significantly influenced by anatomical differences between speakers, such as vocal tract length (Adank, Smits, and van Hout 2004). Normalization is a way of minimizing the influence of these anatomical differences. There is some collateral damage in normalization to the level of sociolinguistically-based variation in the data, but not enough to cause serious problems with the interpretation of sociolinguistic data (pp. 3099-3100). Normalization methods can be vowel-intrinsic or vowel-extrinsic. Vowel-intrinsic methods, such as Bark and MEL transforms (Traunmuller 1990;

Stevens and Volkmann 1940) are usually based on a nonlinear scaling of the raw data, and they do not require any data outside of what can be found within the vowel. Vowel-extrinsic methods use data from across multiple vowels in order to establish baseline characteristics about speakers. As one example, Lobanov (1971) normalized formant values by subtracting the mean formant value across all monophthongal values for a given speaker from the non-normalized formant value, and dividing the result by the standard deviation of all monophthongal vowels together. Methods can also be formant-intrinsic (relying only on single formant values to normalize those formants) or formant-extrinsic (relying on multiple other formants to normalize one formant value). Adank et al. (2004) found that vowel-extrinsic and formant-intrinsic methods were more reliable at minimizing biological differences and preserving phonemic and sociolinguistic information. However, vowel-intrinsic methods have their advantages: It does not require a sampling of the entire vowel space—which is especially challenging when working with corpus data. They have the additional advantage of being qualitatively similar to human vowel normalization, which is refined by but not dependent upon sampling the whole vowel space (Clopper 2009). For these reasons, we chose to use a vowel-intrinsic method (Bark) with our own data set before investigation.

To this end, we adapted the vowel-intrinsic norm.bark formulae from the vowels package in R (Kendall and Thomas 2014). Given a token, its F1/F2/F3 values were first converted to their Bark-scale Z1/Z2/Z3 using Traunmuller's equation (1997):

$$Z_i = \frac{26.81}{(1 + \frac{1960}{F_i})} - 0.53$$

Adank et al. (2004) suggest that Bark conversion is the only step in normalization, but following the advice of Kendall and Thomas (2014), we then calculated Z3-Z1 as a model for each token's height (in place of F1), and Z3-Z2 to model advancement (in place of F2-F1). Unless otherwise noted, height and advancement values are always reported in Bark, and graphs involving formant values use the normalized vowels. One important disadvantage of this method is that normalizations of F1 and F2 into Bark are now partially dependent on the value of F3, for which (from anecdotal observation) Praat's formant tracking is somewhat less reliable.

Following boundary adjustment and automatic measurement, we determined two measures of diphthongization: one perceptual, one acoustic. For the perceptual measure, we annotated the status of each token subject to diphthongization based on an impressionistic index. The annotator was the author—a non-native but competent speaker of French who learned the language exclusively in Québec, in the classroom and workplace (L3, age 18). Tokens were thrown out if their formant values were clearly being mistakenly characterized by the measurements built into Praat. However, uncertainty about a perceptual assessment was not a criterion for throwing out a token; *every* good token was marked on this scale:

(0): non-diphthongized

(1): diphthongized

(2): not diphthongized, but the /ɛ:/ token in question showed a mutated value, generally lowered and slightly backed to [a], corresponding to the

onset of a typical /ɛ:/ in Québec French. This phenomenon occurred rarely (2.9% of tokens), and almost exclusively before /ʀ/ (96% of these tokens occurred before /ʀ/, when /ʀ/ tokens were 74.4% of the data). It occurred more often in rapid speech, and more commonly with female speakers; it is unclear what the relationship is between this process and that of diphthongization.

To correspond with the perceptual index, an acoustic measure of diphthongization was also used for analysis, using the Euclidean distance between Bark-normalized onset and glide points in Z3-Z1 × Z3-Z2 space, as was used in (for example) Mackenzie and Sankoff (2010), to measure the same variable being studied here:

$$dist = \sqrt{((Z3-Z2)_{gl} - (Z3-Z2)_{ons})^2 + ((Z3-Z1)_{gl} - (Z3-Z1)_{ons})^2}$$

The distribution of distances in the data corresponds well to a normal distribution after log-transformation by the formula $dist_{xfm}=\log_{10}(d+0.2)$; these values were used in all graphs and regression models, unless otherwise noted. Besides these differences, in addition to F1 and F2, the perception of vowels can be influenced by the values of higher formants such as F3 (Johnson, 1-3); the normalization algorithm accounts for F3, though not for F0 or consonant factors, which are also relevant to vowel perception.

"Attention to speech" in the Labovian sense was easily monitored; each sound file had previously been tagged by Milne (2014) for one of two connected styles: reading (from prepared remarks, 42.1% of total tokens), and spontaneous speech (57.9%). In line with Labov (1972, ch. 3), the reading-style speech was considered to be higher-attention, which is hypothesized to show a greater incidence of nonstandard speech variables, as well as slower speaking rates, less reduction of vowels, and certain language-specific phenomena. Milne (2014), investigating some of these phenomena in this dataset, found significant style-shifting effects in the rate of consonant cluster simplification; it increased from 57.6% in prepared speech to 76.6% in spontaneous speech (pp. 81-82). He also found a lower rate of schwa epenthesis in spontaneous speech, a mere 6.1% in the ANQ corpus, compared to 17.6% in prepared speech (p. 97), and a faster speaking rate in spontaneous speech, 13.26 segments per second in spontaneous Québec speech vs. 12.40 in reading style (p. 83). Since consonant cluster reduction is a well-studied stratified variable in French, its variability between speech styles indicates that the distinction between reading and spontaneous styles in this data set is a viable representation of the attention-to-speech continua in the Labovian sense.

Since there is no sanctioned clear speech task within this dataset, "clear speech" is measured indirectly, through acoustic features known to correlate with clear speech. Smiljanic and Bradlow (2009) identify some of these factors: speaking rate, pitch range, SPL, and size of a speaker's vowel space during a given sound file. Speaking rate, in this case, was measured in segments per second by word--more precisely, the number of segments in the token's word divided by the duration of the word as determined by the force-aligner. In this paper, we will be investigating speaking rate, frequency, neighborhood density, and segment duration as known

correlates of clear speech. We will not investigate them here, but as seen in Smiljanic and Bradlow (2009), language-specific measures can be as important to the characterization of clear speech as language-universal measures. Two useful language-specific measures of hyperarticulation could be the aforementioned consonant cluster reduction and schwa epenthesis processes. These are both phonetically natural processes, in line with Zwicky (1972). Given the general tendency towards CV syllables--consonant cluster reduction turns CC codas into C codas, while schwa epenthesis turns one CVC syllable into two CV syllables (or breaks up consonant sequences between syllables)--both of these processes can be seen as changing the orientation of speech. That is, the presence of consonant cluster reduction indicates speaker-oriented speech (because it is an articulatory simplification as compared to non-reduced clusters) and the presence of schwa-epenthesis indicates more listener-oriented speech (because it is an articulatory complication).

3. Results

3.1 Exploratory data analysis

We manually trimmed the script's automatically-isolated tokens, to remove words which the script falsely identified as subject to diphthongization; it is possible that there were some false negatives, where the script failed to pick out some tokens which would have been subject to diphthongization. In the end, we were left with 1771 usable tokens of speech from 61 speakers. Of those speakers, the male/female breakdown was 41/20. However, the number of tokens per speaker was not uniformly distributed, with only 20 speakers having over 25 tokens of /ɛ:/. A similar but not-identical subset of the speakers, 21 total, had at least 5 tokens of inherently-long /ɛ:/ and 5 of allophonically-long /ɛ:/. Overall, speakers ranged between 1 and 100 tokens, except for one speaker (#92) with 124 tokens and one (#23) with 299. See Figure 1 for the distribution of number of tokens.

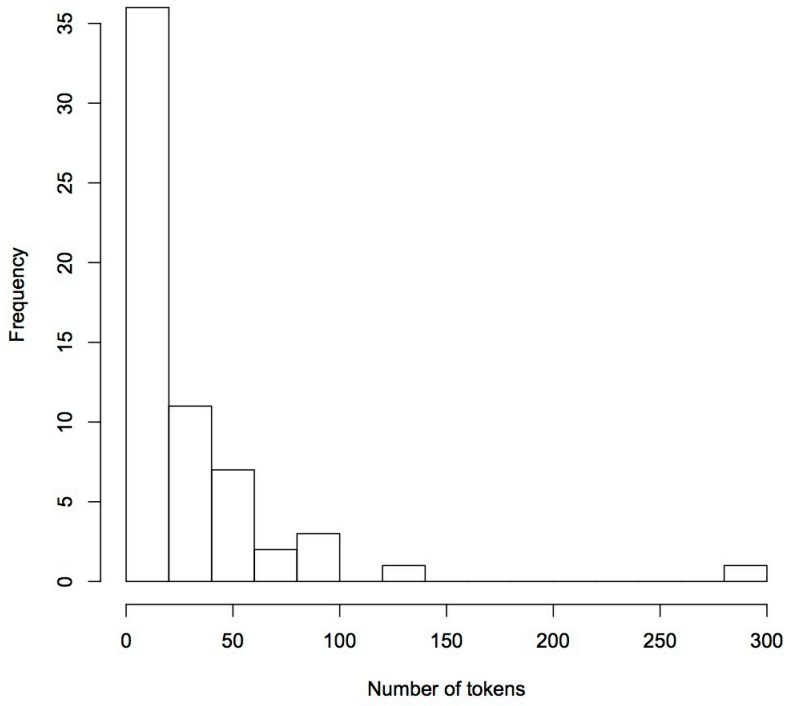

Figure 1. Histogram of number of tokens by speaker

Table 1. Diphthong status of tokens (n = 1771)

Non-diphthongized	*Diphthongized*	*Mutated*
1244 (70.3%)	477 (26.8%)	50 (2.9%)

In order to get a rough sense of which factors are important in the diphthongization of /ɛ:/, we first undertook some exploratory data analysis. The goal was to identify factors which may have an effect on diphthongization. Using factors found by our analysis, plus those suggested by prior research (even if they did not show visible effects in the exploratory data), we would later be able to build regression models to determine the significance of any observed differences. Impressionistic coding of the data reveals that diphthongization of long /ɛ:/ is still reasonably prevalent in Québec French as a whole, as of 2011. Of all tokens, 26.8% showed diphthongization, and a further 2.9% of tokens showing a non-diphthongized but mutated value, closer to [a] than [ɛ]; this mutation was more common in females than males (4.5% vs. 2.2%), and affected tokens before /ʁ/ almost exclusively (96%). As will be seen later, these tokens were found more often in more rapid speech contexts. On its face, the diphthongization rate has declined somewhat since the Santerre and Millo study, which found diphthongization of /ɛ:/ (notated there as /ɜ/) to be realized at an overall rate of 36% (p. 176). However, the decline is not precipitous, as one might have predicted from Santerre's finding that diphthongization was becoming checked among the Québec middle class well over 30 years prior to the collection of

the ANQ data. Additionally, we can neutralize the possible effects of following context by calculating the diphthongization rate for all eight following contexts and taking the mean of those measures. A full breakdown can be seen in Table 5; the average rate across contexts is 43%. Our finding may be in line with Mackenzie and Sankoff (2010) as well, which found an overall decrease between 1971 and 1984 in diphthongization of long vowels, but no significant decrease with /ɛ:/.

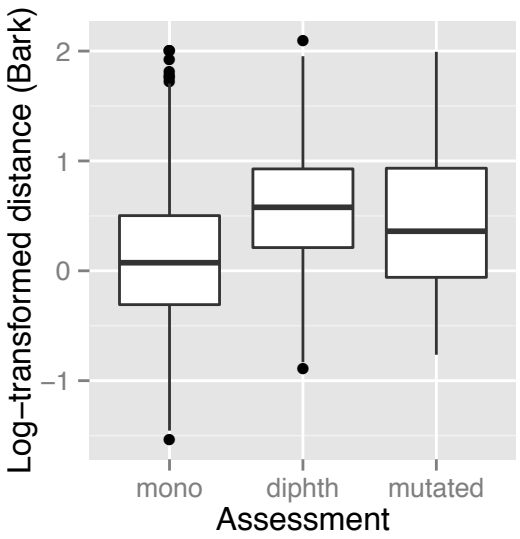

Figure 2. Boxplot of log-transformed distance as a function of diphthongization assessment

Since we are using two separate measures of diphthongization, it is useful to investigate how they correlate with each other, to ensure that they are measuring the same phenomenon. The above boxplot plots the distances between the onset and glide in normalized Bark for the two major assessments, monophthongized and diphthongized (discarding the "mutated" tokens as described in the Methods section). There is wide variability and some overlap between the two assessments, suggesting that diphthongization is a continuous variable—there are tokens which are not clearly one or the other. The median for monophthongized tokens is very close to zero, where the diphthongized tokens are centered around just over 0.5 transformed Bark. Overall, it can be seen that there *is* a trustworthy correlation between the annotator's assessments and the true value of the distance, but distance in F1 x F2 space may not be the only factor in determining the assessment of diphthongization, and even if it is, the automatic measurements may not always be faithfully capturing the transitions in distance which are responsible for the perception of a given token as a diphthong.

The two methods of measurement are distinct but complementary. The distance measure is a singular, reliable factor for degree of diphthongization, which had been used previously for the same purpose by Mackenzie and Sankoff (2010) among others; the assessment takes into account more of the nuances that cue diphthongization, but is less likely to be trustworthy. However, the evidence for distance as a useful correlate of diphthongization is strong enough

that we will be analyzing both the acoustic and perceptual measures throughout the rest of the paper.

3.2 Factors associated with attention to speech

As discussed above, speech style, as an indicator of attention to speech, would be expected to affect diphthongization. Interestingly, there were no large differences by reading vs. spontaneous style in the aggregate data set, with a diphthongization rate of 27.1% in spontaneous style and 26.6% in reading style, which suggests that unlike consonant cluster reduction, the diphthongization of long vowels is not subject to strong attention effects. However, we can look more closely at interspeaker variability in any desired variable when faceting the data by speaker. This is particularly desirable for style-shifting, which is known not to affect different speakers at equal rates, especially in cases of hypercorrection by the upper-middle class (Wardhaugh 2002, p. 167). Furthermore, the number of tokens for each speaker is nowhere close to equal. If we do this for Euclidean distance, it can then be seen that for some speakers, there are noticeable differences in the amount of diphthongization by style. Among the 13 speakers who had over 25 tokens total and at least one in each style, 7 speakers (5, 15, 31, 40, 55, 66, and 67) seemed to show the pattern that one would expect under the Labovian model of attention to speech, where the nonstandard variable (greater diphthongization, i.e. greater distance) was seen more in the more casual spontaneous style. Four (9, 12, 23, and 92) showed very close to no style shifting, while two (62, and 105) showed the opposite trend, none very strongly. For most speakers, there was a wide degree of variability in all categories, which partially explains why the aggregate data showed no especially strong trends in style. From our investigation of this inter-speaker variability, we suspect that it might be possible to see a significant effect in the expected direction for certain speakers.

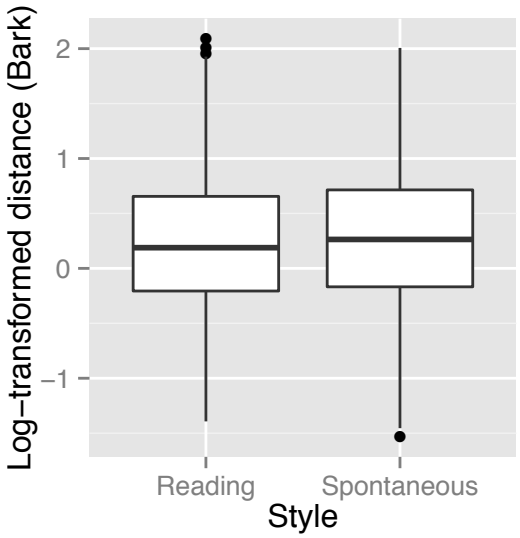

Figure 3. Boxplot of log-transformed distance as a function of attention to speech, aggregate

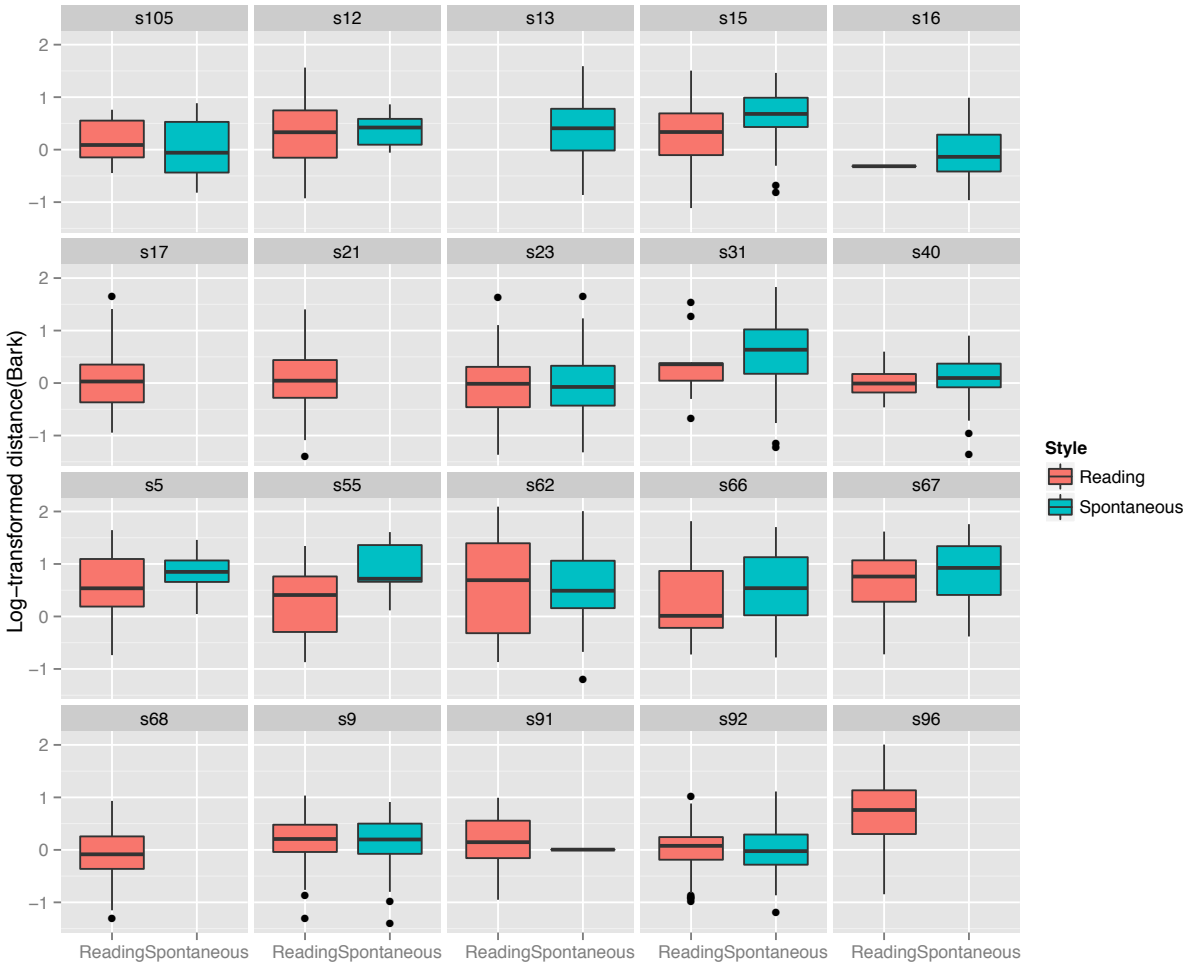

Figure 4. Boxplot of log-transformed distance as a function of attention, for each speaker with 25+ tokens

Table 2. Breakdown of perceptual diphthongization rate by speech style

	Spontaneous speech (N = 746)	*Reading style (N = 1025)*
monoph.	517 (69.3%)	728 (71.0%)
diph.	202 (27.1%)	273 (26.6%)
mutated	27 (3.6%)	24 (2.3%)

3.3 Factors associated with clear speech:

Having examined speaking style--a factor associated with attention to speech--we can now turn to some of the measures correlated with the clear-speech continuum as described in previous literature (Gahl et al. 2011, Smiljanic and Bradlow 2009, etc.), to determine whether these measures affect diphthongization (and, indirectly, the perception thereof), in the predicted way.

3.3.1 Speaking rate

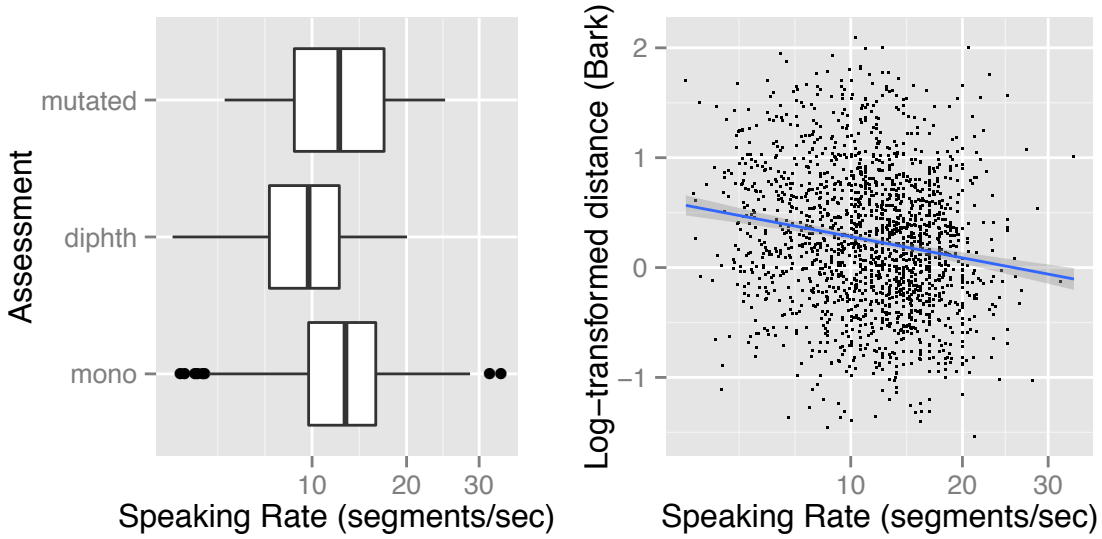

Figure 5. Scatterplot of distance (L) and boxplot of assessment (R) by speaking rate

Interestingly, when examining some of the aforementioned factors correlated with clear speech, the opposite pattern can be seen, and more strongly (Fig. 5). We had also been investigating measures of speech clarity in the data, and one important measure of this is the speaking rate. In the above graphs, the speaker's speaking rate is plotted against the transformed normalized distance between onset and glide in Z3-Z1 × Z3-Z2 space. It can be seen in the aggregate graph that as one approaches the faster speaking rates, the smoothed best-fit line approaches a distance of 0, overshooting to the negatives in the fastest sections of speech; note that the negative distance is simply a consequence of the log transformation, and not related to any measure of direction. Except for speaker 96, who shows a slightly reversed pattern (with high distance measurements throughout), and speakers 66 and 105, who show a nearly flat profile, the downward trend holds to some extent for every speaker in the data set with 25 or more tokens. Additionally, as seen in the middle graph, *none* of the fastest tokens--those with a speaking rate of over 20 segments per second--were rated as being diphthongized; this is why it was particularly important to look at the acoustic measure as well as the perceptual measure when considering speaking rate.

The correlation between speaking rate and distance appears weak, with a great degree of variability in diphthong distance even at faster speaking rates. This finding is largely in keeping with the observation from Lindblom (1990) that, although there are strong trends towards undershoot of phonetic targets in rapid speech, "speakers have a choice" (p. 414) to hyperarticulate (or, more precisely, avoid hypospeech) even at higher speaking rates. In other words, despite trends, the effects of hypospeech cannot be expected to be seen in every individual token. This trend can be seen clearly in the perceptual data as well. The annotator's perceptions of diphthongization show the same trend borne out in the acoustic data; however, there was a large amount of overlap--as can be seen in the graph, almost half of the respective IQRs of the speaking rates for tokens assessed as 0 or 1 overlap. As mentioned above, the mutated-value tokens (in blue) patterned more closely with the non-diphthongized tokens; they were more likely to be seen in faster speech. Overall, diphthongization decreases noticeably at faster speaking rates, in both the acoustic and perceptual data, especially the latter.

3.3.2 Segment duration

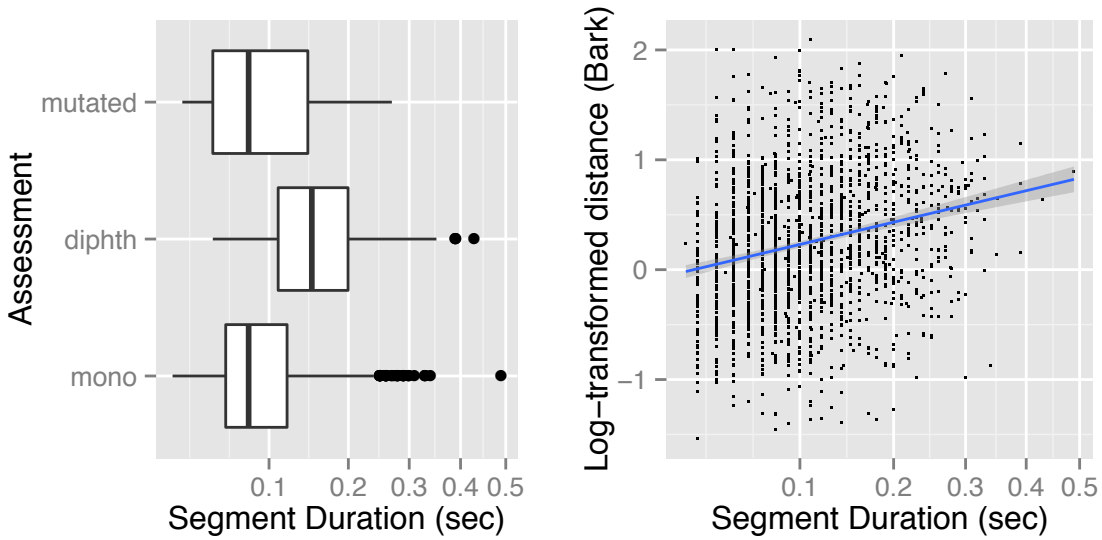

Figure 6. Boxplot of assessment by duration (L) and scatterplot of distance by duration

Segment duration patterned similarly to speaking rate (but in the opposite direction), and thus appeared to be correlated with both measures of diphthongization. Longer segments tended to show greater distances, although short segments could show large distance measurements as well, and long segments could show relatively little movement. Tokens which were perceived as diphthongized had longer durations on average than those perceived as monophthongized or mutated.

3.3.3 Frequency and neighborhood density

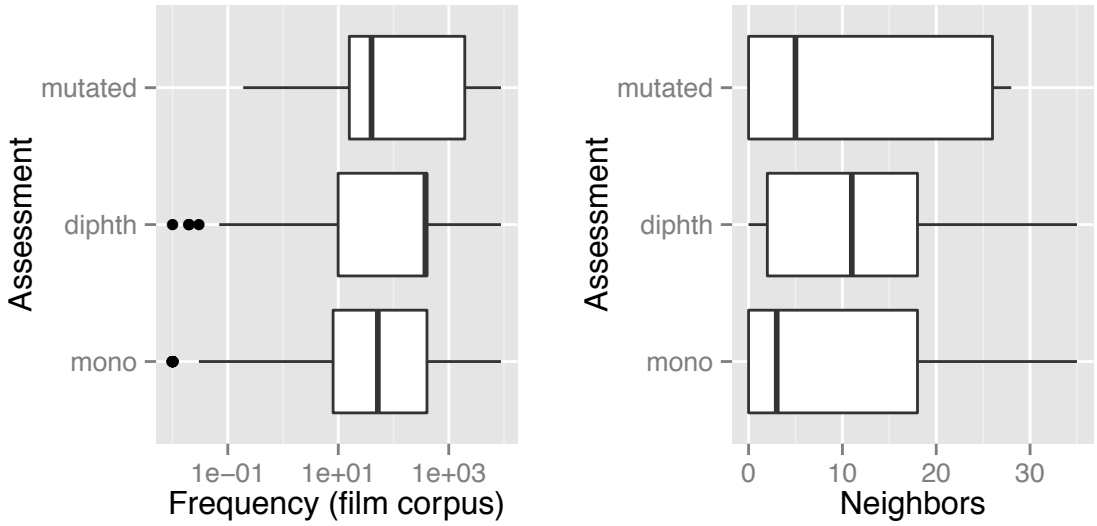

Figure 7. Boxplots of assessment as a function of frequency, log scale (L) and neighborhood density (R)

As per Gahl, Yao, and Johnson (2011), we investigated the relationship of diphthongization to the frequency and neighborhood density of the words where it occurs. In this study, we used a log-transformed frequency measure from a film corpus, as published in Lexique (New and Pallier, 2001), and the number of phonological neighbors within the Lexique data. Though there was significant variability, the median frequency and neighborhood density of the diphthongized tokens was noticeably higher than in the non-diphthongized tokens; again, the "mutated" tokens patterned here with the monophthongs. These are intuitive results, which line up with previous work (the ND finding agreeing with the studies cited in Gahl et al. but not Gahl et al. itself). Diphthongization was more likely in higher-frequency words and words with more neighbors. This could be because since such words are more susceptible to confusion--as outlined by Lindblom (1990)--they require more hyperarticulation to achieve sufficient discriminability. However, high-frequency words also tend to be realized at shorter durations (Gahl et al.), and shorter words are less conducive to diphthongization. Furthermore, we cannot conclude that diphthongization is used strategically to distinguish between long and short /ɛ(:)/, since the phonemic status of long /ɛ:/ is marginal at best. The results for neighborhood density should also be handled carefully; the number of neighbors was calculated without regard for any (quasi-phonemic) distinction between long /ɛ:/ and short /ɛ/, since Lexique does not include those distinctions in their pronunciation guides.

3.4 Miscellaneous factors of interest

The following factors are not necessarily associated with our continua of interest—attention to speech and clear speech—but for various reasons, they were hypothesized to have a sociolinguistic or purely phonological effect on the rate of diphthongization. We made sure to explore these in order to ensure that we were not missing anything which contributed significantly to diphthongization.

3.4.1 Sex

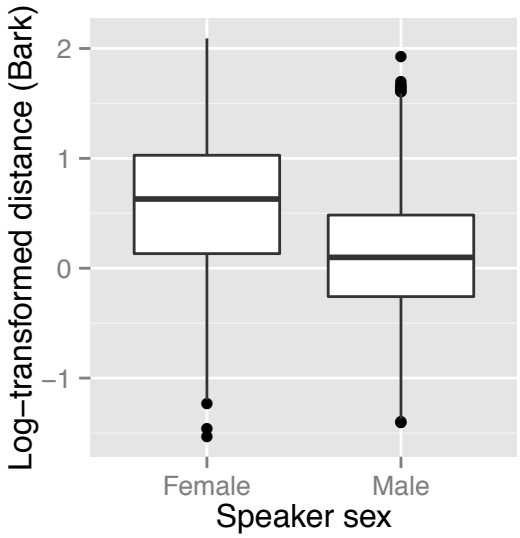

Figure 8. Boxplot of log-transformed distance as a function of speaker sex

Male speakers were somewhat more likely to show diphthongization than female speakers. This finding lines up with the general observation that female speakers are more likely to realize the prestige variant than male speakers in stable stratified variables, and also to lead changes, such as the slow decline of nonrhoticity in NYC English or—in our case—of diphthongization in Québec French (Wardhaugh 2002). The importance of the modest differences observed in our data is heightened by the fact that females tend to have slower speaking rates than males (12.081 segments/second for females in our data, 12.126 for males, $p < 0.05$), so under the clear-speech model, they would be expected to realize *more* diphthongization, not less.

Table 3. Number and percentage of tokens diphthongized by speaker sex

Total *(61 speakers, 1771 tokens)*	*Male* *(41 speakers, 1240 tokens)*	*Female* *(20 speakers, 531 tokens)*
477 (26.8%)	361 (29.1%)	116 (21.8%)

In fact, we do (in a way) see that women diphthongize more than men. When looking at the *acoustic* measures of diphthongization, the opposite pattern emerges.

In both assessment categories ("mutated" tokens excluded), females show greater vowel movement in F1 x F2 space from the 25% point to the 75% point in the vowel. We will later use regression models to determine the significance of the observed sex effects, and discuss possible reasons for these findings afterwards, if the effects are found to be significant.

3.4.2 Word class and context

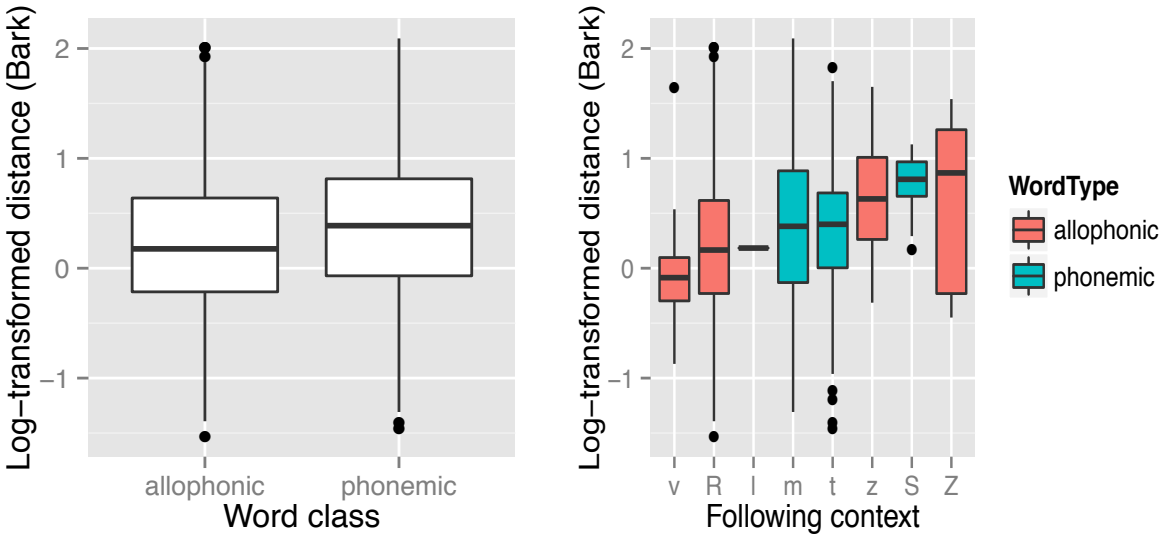

Figure 9. Boxplots of log-transformed distance as a function of word class (L) and following context (R)

Table 4. Breakdown of perceptual assessment by word class

	Type 0: allophonic /ʁ z ʒ v/ (N = 1404, 79.3%)	*Type 1: Historically long <ê> (N = 367, 20.7%)*
monoph.	1086 (78%)	159 (43%)
diph.	269 (19%)	206 (56%)
mutated	49 (3%)	2 (1%)

Table 5. Diphthongization rate by immediate context (bold = phonemic class; plain = allophonic class)

	l	***m***	***ʃ***	***t***	*ʁ*	*v*	*z*	*ʒ*
n	**1**	**4**	**2**	**12**	181	10	9	3
monoph.	**1** ***100%***	**70** ***35%***	**1** ***9%***	**84** ***56%***	1044 *79%*	21 *91%*	15 *27%*	4 *57%*
diph.	**0** ***0%***	**127** ***64%***	**10** ***91%***	**67** ***44%***	226 *17%*	2 *9%*	40 *73%*	3 *43%*
mut.	**0** ***0%***	**2** ***1%***	**0** ***0%***	**0** ***0%***	49 *4%*	0 *0%*	0 *0%*	0 *0%*

At first, there appeared to be a very large effect of word class. Recall that among allophonic tokens, only 19% were diphthongized, mostly in the context of /ʁ/ (*financière, ministère,* etc.); however, when looking at historically-long tokens (*même, pêche, enquête*), the

diphthongization rate rose to 56%. This difference is also seen in the plots of distance vs. word class (Fig. 9). Because of the wide variability among all tokens, a great deal of overlap is visible, but the mean distance for allophonic tokens (0.215) is still lower than for phonemic tokens (0.364).

However, this effect is likely to be at least partially illusory; it may be more accurately explained as an effect of the following phonological context. When looking at the diphthongization rates of /ɛ:/ by specific context, a slightly different pattern emerges.

As can be seen in Table 5, diphthongization rates are widely variable even within the "allophonic"/"phonemic" groups. Notably, out of all of the contexts with more than 20 tokens, before /z/ is actually the most likely to diphthongize, even though it's in the allophonic group, which has lower diphthongization rates as a whole than the phonemic group. Breaking down diphthongization rates by specific context, rather than the allophonic/phonemic distinction, shows that the differences in diphthongization between the two larger categories are illusory. The most likely cause is that tokens of /ɛ/ before /ʁ/ have a low rate of diphthongization, and since they make up 74.6% of the total tokens, they bring down the average of the group as a whole. There appears to be a phonetic component to the distribution of diphthongization, as vowels appearing before a coronal fricative often showed higher movement than in other environments. However, it should be noted that several of the contexts have not only few tokens but few different words. For example, the data on /ʃ/ consist of just two words--two instances of *empêche* and nine of *pêche*. Save two tokens of *extrême*, *every* instance of a long vowel in context /m/ is *même*. So although there is an apparent effect of context and particularly of place of articulation, the unique behavior of individual lexical items cannot quite be discounted. We will be able to treat words as a random effect in the regression models, to ensure that this does not confound our results.

The exploratory data analysis, as a first step, suggests that many of the trends we had anticipated will be borne out. Though attention to speech was not found to be radically important in the aggregate data, it appeared to have effects in the expected direction for some individual speakers, with greater diphthongization in more casual styles, suggesting that reading and spontaneous styles should be explored further. We found several tentative effects of context, such as greater diphthongization in historically-long tokens, although it is uncertain whether they are motivated by some underlying phonological principle, by differences between allophonic and historically-long tokens, or simply by the uneven numbers of tokens for each following context. One of the most interesting findings has been the differing effect of sex on the two measures: though female speakers were found to have lower diphthongization rates overall by the perceptual measure, they had higher average formant movement across all tokens of /ɛ:/. It is uncertain whether these effects are truly significant, and it will be important to explore them further in the regression analysis.

3.5 Regression analysis

The exploratory data gave us a good understanding of the basic factors affecting diphthongization in Québec French, but because this study is multivariate, these simple models

do not give us a good picture of how these factors work *together* in order to influence the dependent variable. In order to test the general trends seen in the exploratory data and answer the initial research questions, we later investigated some regression models for the independent variables explored above.

Regressions are designed to model the dependent variable (in our case, either the distance measure or the perceptual assessment of diphthongization), as a function of a number of independent variables chosen during the exploratory data analysis (speaking rate, style, sex, etc.). Two models were necessary to confirm the effects found in the exploratory data: a logistic regression model, for the perceptual data (assessment of diphthongized/non-diphthongized, mutated tokens discarded); and a linear regression model, for the continuous acoustic data (Euclidean distance).

The table below summarizes the logistic regression model for the binary variable of diphthongized/not diphthongized, read as a numeric where diphthongized = 1. In other words, the "estimate" column is measuring the log-transformed odds that a given token is diphthongized. The estimate of the intercept is seen to correspond very closely to the log-transformed odds of a token being diphthongized; $10^{-0.6101}=0.2454$, which is close to the measured probability of 0.261. For our purposes, significance will be measured thus: $p < 0.05$ is significant; $0.05 < p < 0.1$ is marginally significant; $p > 0.1$ is not significant.

One should note that the regression models use two separate measures of phonological context. Given the results of the exploratory data analysis, we suspected that the apparent effect of phonemic vs. allophonic tokens was likely to be an illusion brought on by the unwarranted clustering of contexts which in reality behaved very differently. Therefore, we used a three-level Helmert coding system to make two binary comparisons: the first between tokens before /ʁ/ and other allophonic tokens (ignoring the phonemic tokens), and the second between word classes, to confirm whether clustering the tokens into fewer but larger groups results gives us significant inter-group differences in diphthongization.

The fixed effects being studied in the regression models were determined by our observations about various factors in the exploratory data analysis; they are as follows:

- *Speaking rate:* Number of segments in token's word divided by duration of token's word, square-root transformed
- *Segment duration:* Duration of token segment. Investigated in the exploratory data but *excluded from the regression models*, because of its functional similarity to speaking rate; additionally including both duration and speaking rate results in a convergence failure in the linear model, and renders speaking rate insignificant in the logistic model.
- *Neighborhood density:* Number of phonological neighbors of the word (edit distance of 1), as per Lexique, log-transformed
- *Frequency:* Frequency of the word in a film corpus, as per Lexique, log-transformed
- *Sex:* Female = 1, Male = 0

- *Style:* Reading = 1, Spontaneous = 0
- *R vs. other allophonic:* Tokens before /ʀ/ vs. tokens before /z ʒ v/ (/ʀ/ = -1 and other allophonic = 1)
- *Word class*: Tokens before /ʀ z ʒ v/ vs. all other contexts (/ʀ/ = -1, allophonic = -1 phonemic = 2).

Out of concern that spurious or idiosyncratic variation within the data set would unduly impact our measures of significance, we implemented random effects into our regression models. We identified the following random effects for each model:

- *Word*: The word where the token appears
- *Speaker*: The speaker realizing the token

In particular, we were uncertain whether diphthongization could be subject to lexical effects of individual words (apart from the effects of word class or context), and given the large number of distinct words, it was simplest to treat this variability as random. Similarly, in our exploratory analysis, we saw good reason to believe that there was inter-speaker variability. Although there could potentially be identifiable correlates of this variability (such as speaker age, region, or sex), these correlates were beyond the scope of this study, so we treated the variability as random. We also used random slopes, to account for the possibility that certain fixed effects had some random variation introduced by the random effects. In the both models, we allowed random slopes as follows: speaking rate, reading style, and context (R vs. other allophonic) had random slopes by speaker, and speaking rate and sex had random slopes by word. In the logistic model, we additionally had speaker as a random intercept, to account for interspeaker variability in diphthongization rates as seen in Figure 4.

Table 6. Fixed effect estimates for logistic regression on perceptual criterion (diphthongized = 1), with standard error, z value, p value, significances

	Estimate	*SE*	*z value*	*p value*	*Significant?*
(Intercept)	-1.13497	0.27761	-4.088	p < 0.0001	yes
Speaking rate	-1.55055	0.17862	-8.681	p < 0.0001	yes
Neighborhood density	0.05412	0.12077	0.448	0.654061	no
Frequency	-0.11105	0.06191	-1.794	0.07284	marginal
Sex	-0.36867	0.20443	-1.803	0.071327	marginal
Style	-0.37001	0.10211	-3.624	0.00029	yes
Context: R vs. other	0.84679	0.25479	3.323	0.000889	yes
Word class	0.34816	0.16334	2.131	0.033049	yes

Table 7. Fixed effects for linear regression on acoustic criterion (log(distance) + 0.2), with standard error, t value, p value, significances

	Estimate	*SE*	*t value*	*p value*	*Significant?*
(Intercept)	0.423097	0.04485	9.434	p < 0.0001	yes
Speaking rate	-0.112833	0.032472	-3.475	0.0015	yes
Neighborhood density	-0.009454	0.018143	-0.521	0.6043	no
Frequency	-0.007785	0.009066	-0.859	0.3970	no
Sex	0.206553	0.029317	7.045	p < 0.0001	yes
Style	-0.043898	0.017369	-2.527	0.0272	yes
Context: R vs. other	0.083631	0.044579	1.876	0.0791	marginal
Word class	0.010443	0.027992	0.373	0.7118	no

Given the results of these models, we will have to interpret each fixed effect individually for three key features: the *direction* of the effect, the *magnitude* of the effect, and the *significance* of the effect. The sign of the estimate for a given effect determines the direction, so in the linear regression on Euclidean distance, a negative sign in the estimate would indicate a negative relationship, and a positive sign would indicate a positive relationship. As one example, as speaking rate decreases, distance increases. The magnitude of the effect is easily interpreted from the *t*-value (in the linear regression) or *z*-value (in the logistic regression), with larger values indicating larger magnitudes, and the sign of the value corresponding to that of the estimate. Finally, the significance is determined by the *p*-value.[4] There is no universally-accepted method to interpret *p*-values,[5] but here we will use the common convention that a *p*-value of less than 0.05 is statistically significant, a *p*-value between 0.05 and 0.1 is marginally significant (resting partially on the assumption that a larger data set might give us a more reliable significance), and a *p*-value above 0.1 is not significant.

3.5.1 Speaking rate

As suspected, speaking rate has a significant negative effect for both models; as speaking rate increases, diphthongization and distance both decrease (diphthongization: beta = -1.55055, z

[4] In the simplest terms, the *p*-value represents the probability that, assuming that the null hypothesis is true, the *t*- or *z*- value would be as extreme in its distance from 0 as it is. The null hypothesis for a given fixed effect is that it does not affect the dependent variable.

[5] In fact, the importance of significance tests is widely debated, and they are somewhat subject to abuse in interpreting statistics; see Parkhurst (1997) for a collection of opinions on the matter.

= -8.681, p < 0.0001; distance: beta = -0.113, t = -3.475, p = 0.0015). It has the largest of any fixed effect in the logistic model, and the second-largest in the linear model.

3.5.2 Segment duration

Segment duration was excluded from our models, but when it was included in the logistic model, it became an enormously powerful predictor of diphthongization--significant and large in magnitude. However, this is misleading, because diphthongized vowels are *inherently* longer on average, independent of the experimental conditions.

3.5.3 Neighborhood density

The number of phonological neighbors of a word is not a significant predictor of diphthongization (diphthongization: beta = 0.05412, z = 0.448, p = 0.654061; distance: beta = -0.009454, t = -0.521, p = 0.6043). As mentioned above and in Gahl, Yao, and Johnson (2011), previous work found significant effects of neighborhood density on hyperarticulation--with more dense words hypoarticulated (Gahl) or hyperarticulated (elsewhere). However, given the highly inconclusive distribution seen in our exploratory data, it is unsurprising that density turned out to be insignificant.

3.5.4 Frequency

Frequency was found to have a negative, marginally-significant effect on the perceptual model (more frequent words are less likely to diphthongize: beta = -0.11105, z = -1.794, p = 0.07284). It had no significant effect on the distance model (p = 0.3970).However, the marginally-significant finding in the perceptual model *contradicts* the finding in the exploratory data analysis, where diphthongized tokens had a higher median frequency (Fig. 7). Given the number of niche words in the data set (and the near-complete lack of very common, closed-class words), the artificially limited sampling (only the small subset of possible words which contain a given token), and their very unusual distribution, it is likely that the corpus simply did not provide enough data to see a significant, fully trustworthy frequency effect.

3.5.5 Sex

The unusual stratification by speaker sex seen above in the exploratory data, where female speakers diphthongize less frequently than males, but show greater average movement in their tokens, is borne out in the regression models, and found to be robustly significant in the acoustic data (beta = 0.206553, t = 7.045, p < 0.0001), but only marginally significant in the perceptual data (beta = -0.36867, z = -1.803, p = 0.071327). Given the ballpark similarity in diphthongization rates (29.1% male vs. 21.8% female) and the smaller number of female tokens, it is unsurprising that this effect falls just above the 0.05 threshold. Much more jarringly, in the distance models, it is actually the most significant effect (largest absolute t value), even larger than speaking rate. It should be noted that sex affects speaking rate, however. In the aggregate ANQ data, male speakers have a mean speaking rate of 12.126 segments per second, and females have a mean speaking rate of 12.081 segments per second; this difference is very small but marginally significant, in line with the aforementioned findings that females speak more slowly

than males. In any case, the finding of sex as significant in a regression indicates that it is significant *independently* of speaking rate; female speakers have less likely diphthongization than males but more radical vowel movement, at significant levels for both. We will discuss male-female differences in greater detail in the discussion section.

3.5.6 Style

Reading style makes diphthongization significantly less likely, and distances smaller (diphthongization: beta = -0.37001, z = -3.624, p = 0.00029; distance: beta = -0.043898, t = -2.527, p = 0.0272). This finding is important, because it demonstrates that more formal speech is not necessarily clearer speech; if it were, we would see increased diphthongization in more formal styles, because clear speech shows more diphthongization than citation speech. Furthermore, it reinforces the status of diphthongization as a vernacular variant, because it is seen to be less commonly-realized in more formal speech. Though not all speakers showed the expected style-shifting in the exploratory data, *none* of them showed style-shifting in the opposite direction. The regression models confirm that *overall*, despite individual variability, style does have the expected effect on both diphthongization and distance. Potential reasons for the individual variability in style-shifting will be discussed below.

3.5.7 Context (/ʁ/ vs. /z ʒ v/)

When comparing tokens before /ʁ/ to tokens before /z ʒ v/ (ignoring phonemic tokens), the latter category can be seen to be more likely to show diphthongization and longer distances, at marginal or fully significant levels (diphthongization: beta = 0.84679, z = 3.323, p = 0.000889; distance: beta = 0.083631, t = 1.876, p = 0.0791). This lines up with the fact that 47% of tokens before /z ʒ v/ were diphthongized (despite the low rates for /v/), versus only 17% of tokens before /ʁ/. It is uncertain if there are articulatory justifications for contextual distinctions, but /z ʒ/ are both coronal fricatives with medium to high diphthongization rates, and the other coronal fricative studied—/ʃ/—also has a high rate. /v/ patterns differently, with a low diphthongization rate, although it is similarly situated at an anterior place of articulation. /ʁ/ is a dorsal trill, two features which it does not share with any other contexts observed. Since coronal fricatives are realized with the apex of the tongue in a higher position (closer to [i] than [ɛ]), it is possible that an upglide to [i] in the second element of the diphthong is articulatorily natural when realizing tokens in that context.

3.5.8 Word class (allophonic vs. phonemic)

Phonemic tokens were found more likely to diphthongize at significant levels (beta = 0.34816, z = 2.131, p = 0.033049), but there was no significant effect of word class on distance. There may be some unconscious bias towards perceiving phonemic tokens as diphthongized, maybe because they are orthographically indicated with a uniform letter <ê> rather than a variety of spellings (<è>, <ai>, <ei>, etc.), which gives an obvious visual association with the variable. However, the finding of questionable-but-plausible significance of word class confirms our suspicions that the exploratory data on it failed to account for the disparate contexts and the sizes of each category.

4. Discussion

4.1 Theoretical basis for findings

After having done several investigations into factors correlated with diphthongization of long /ɛ:/ in Québec French, using both impressionistic and acoustic measurements, we have been able to determine which of these factors are actually significant in determining diphthongization. These factors so far fall into two major stylistic categories. First, there are those used most widely in sociolinguistics: sex, attention to speech by style--but style is the primary focus. Second, there are measures related to hyper- and hypospeech: segment duration, speaking rate, word frequency, and neighborhood density.

When investigating a variable like QF diphthongization, these continua are put into conflict. The studied variable is marked, nonstandard, and known to be a more working-class feature, all of which suggest that in higher-attention styles, it would be suppressed. At the same time, it is a realization that (arguably) requires greater articulatory effort than the standard variant, so it may be highlighted in higher-clarity situations. Both of these predictions have been borne out to various degrees so far, the latter more so than the former. We have found that diphthongization is suppressed in more formal speech, but used more often in clearer speech—these relationships are both found in the expected direction, but with the variable in question, the expected directions are *negatively* correlated with each other, where the expected correlation is a positive one. In other words, counterintuitively, more attentive speech is not necessarily clearer speech. It would appear so far that these two stylistic continua, rarely studied together, are distinct from each other, even though they are usually correlated. If this is the case, then there could be such a thing as high-attention, low-clarity speech and vice versa.

Perhaps the finding that these continua are distinct and potentially conflicting is not so surprising, though--there may be different motivations at play for each. Much of the sociolinguistic literature focuses on prestige, and the orderly heterogeneity that speech variables show in various contexts suggest that speakers strive to be *respected* by the listener. However, the clear-speech literature focuses more on speakers' motivation to be *understood*. Thinking of it in those terms, it is easier to see how these differing stylistic continua may not be relevant to each other.

4.2 Explanation of sex differences

The finding on sex differences between male and female speakers gives us some important insight into our original research questions. Perceptually, female speakers diphthongized at a lower rate than males; however, acoustically speaking, female speakers showed *greater* distance on average--both were statistically significant in the models discussed above. Let us first consider two major findings about our two speech dimensions of interest:

1. Females speak more clearly than males, by measures such as speaking rate, vowel space, and pitch range. (e.g. Diehl 1996, Simpson 2009)
2. Given a stable, socially-stratified speech variable, i.e. one which does not represent a change in progress, female speakers are more likely to realize the standard variant. (e.g. Fischer 1958)

Our findings are counterintuitive if considering the attention-to-speech models of style-shifting, because such a model would expect women to diphthongize less than men. The evidence is more equivocal about clear-speech differences between men and women, depending on the study. Sex was not found to be a significant predictor of hyperarticulation in Gahl et al. (2011), both when measuring vowel duration and vowel space dispersion. Byrd (1994) found that, in the TIMIT corpus of English, men spoke 6.2% faster than women, as measured in syllables per second; this difference was significant, albeit small. Numerous studies have found that females have a larger vowel space than males (e.g. Diehl et al. 1996, Simpson 2009), which *is* a marker of clear speech. However, Diehl et al. (1996) note that the reasoning for this dispersion is uncertain. They explore some sociophonetic possibilities, particularly the tendency towards sexual dimorphism. The purported female tendency towards clear speech may be culturally-specific; Arabic-speaking women, for instance, do not show a marked difference in clarity from men (Goldstein 1980, 235 in Diehl et al. 1996). In any case, the characterization of clear speech as feminine could be "more of an effect than a cause" of vowel dispersion (Diehl et al., p. 190). The underlying cause could be biological; given that females have F0 values up to 90% higher than males on average, higher frequency results in poorer resolution of spectral peaks, and thus greater difficulty in distinguishing between vowels, which difficulty is then resolved by expanding the vowel space (Diehl et al.). Clear speech is still clear speech regardless of the reasoning for using it, but the reasoning is important in establishing whether the female tendency towards clear speech is qualitatively different from the female tendency to avoid nonstandard forms. Unless one subscribes to Kroch's (1978) assertion that nonstandard forms are necessarily easier to articulate, the latter is purely sociolinguistic and not a biological tendency.

There are several possible explanations to reconcile these seemingly conflicting findings. First of all, we have already established that Euclidean distance does not seem to create a comprehensive picture of the perception of diphthongization; however, the only other obvious factor which appears to be important to the perception of diphthongization is speaking rate, and female speakers tend to talk at a *slower* rate than males (see Byrd 1994). The difference in speaking rate, if anything, would contribute to an inflated perception of diphthongization in female speech and thus a *narrower* gap between male and female diphthongization rates. The other, more interesting explanation is that female speakers diphthongize less frequently, but when they *do* diphthongize, they do it more radically--so much so that it affects the average Euclidean distance of all of their tokens greatly.

The pattern seen in the data could reflect a resolution to the initial conflict posed by the diphthongization variable. We had been treating diphthongization as an atomic, one-dimensional variable; though we measured it in two ways--perceptual and acoustic--they were essentially variations on a theme. We were uncertain whether diphthongization should be treated as a binary variable--diphthongized or not--or whether it was actually on a spectrum, from zero intra-vowel articulator activity to very large intra-vowel motion. In either case, however, we had not considered that there could be more than one way to modulate the variable. In that way, we expected them to behave very similarly--perhaps one measure would catch a nuance that the

other would fail to notice, but they wouldn't give conflicting pictures of the same variable. In fact, the two measures ended up *representing* the very conflict which they were proposed in order to resolve. The female speakers could be credited here with finding a way to solve the conflict evident between attentive and clear speech, as posed by a variable which does not follow the usual correlation direction between the two: they controlled the social dimension of speech style by modulating the *frequency* of the nonstandard variable, and simultaneously controlled the clarity dimension of their speech by modulating the *degree* of the variable.

4.3 Potential limitations, sociolinguistic attitudes

While our findings here were significant, it is important to also consider the study's limitations. One potential limitation of this study which must be considered is that the social analysis of diphthongization in Québec French may not be as clear-cut as we are assuming. Its stigmatization is inferred from its status as a working-class variable (see Santerre and Millo) and a marked feature of Québec French, the vast majority of whose speakers consider it a distinctive dialect (see Maurais 2008). The style-shifting present in some speakers, and the lower rates of diphthongization seen in female speakers, can both be interpreted as confirmatory evidence of the variable as socially stratified, in the same vein as (for example) (r) in New York City English. Despite the social significance of prestige in speech, there is also a notion of *covert prestige*, where nonstandard forms are embraced by a group in the face of stigmatization, out of a sense of solidarity against the more mainstream prestige paradigm. In a study on working-class dialects in Norwich English, Trudgill (1972) found that working-class men were proud of their nonstandard dialect, despite its poor social standing, because they saw it as a marker of in-group belonging rather than (for example) poor education. It seems plausible that Québec French could be subject to a similar evaluation, vis-à-vis a European or international standard, though the presence of an extrinsic standard is not *necessary* for the stigmatization of a variant. Subjective surveys on conscious attitudes do not trump more (but not fully) objective measures of unconscious biases in speech, but they *can* help us understand the origins of speakers' self-evaluation.

In reality, speakers of Québec French have a diverse variety of attitudes towards their own speech. Though there is widespread agreement (77.6%) that spoken Québec French is a distinctive dialect as compared to European French, and that the differences between the two dialects can create communication difficulties (73.9%) there is not nearly as much agreement about the implications of this distinction (Maurais 2008, p. 19). For example, in 2004, it was found that there is a near-even split between those who claimed that they attempted to modulate their manner of speaking with European francophones (50.6%), and those who speak to Europeans and Québécois the same way (49.4%) (Maurais, p. 21). Though they should not be taken to be directly comparable, since they are measuring different phenomena, these percentages line up closely with the faceted boxplots of style-shifting above, where 7/13 speakers were found to style-shift in the predicted Labovian fashion. It is very possible that, if we were to survey the speakers of the ANQ corpus, we would find that those who style-shifted would also be those who claim to modulate their speech around speakers of other dialects. It is

well-documented (e.g. Labov 1972, ch. 2-3) that upper-middle-class speakers are very attuned to speech stigmatization. We may also find some speakers who claim not to modulate their speech but actually do. However, despite the above evidence of widespread awareness of differences between European and Québec French, it is also uncertain whether European French truly represents the main extrinsic standard for speakers of Québec French, or whether they evaluate their speech based mostly on distinctions found within their own community. Future work with live subjects could consider speakers' backgrounds more closely, and interview them about their linguistic attitudes, to match up individual dialect evaluations with how they actually speak in practice.

Another limitation is the highly subjective nature of the perceptual annotation process. The annotator was working solo on this project, and is a capable and linguistically-attuned but still non-native speaker of Québec French. The formant tracks as measured by Praat were visible throughout the process, which was a tradeoff--the ability to see them could interfere with the perception of a diphthong; on the other hand, it was easier to mark tokens which would pose obvious problems if their formants were measured and automatically entered into the data set. The annotations could of course reflect a confirmation bias, where the trends we hoped to see were magnified in the perception of diphthongization--especially given the number of gray-area tokens. However, there were safeguards against this: The annotation was performed at a very early stage of the project, with few expectations in place, and before the factors to be measured acoustically were even determined; it was also reinforced by an acoustic measurement which reflected the perceptual measurement well. However, it is likely that our results would be more valuable if they were annotated by one or more native speakers of Québec French.

4.4 Questions for future study

Future research on the relationship between variations in articulatory and sociolinguistic factors is potentially a rich field. More research must be done on the distribution of stratified variables: how often does the vernacular variant require more, less, or the same amount of articulatory effort as the standard variant? Additionally, other variables which present the same conflicting picture as Québec French diphthongization--such as the short-a split and cot-caught splits in NYC English--should be investigated, to see if they show the same patterns. Testing these variables in controlled experimental environments could also prove beneficial to our understanding of style-shifting and speech variability. One hypothetical scenario where understanding X further would shed light would be the following: Imagine a middle-aged speaker of Québec French who works in an assisted-living facility. Her clients would largely be local and elderly, and therefore would speak an older form of the same dialect. If a hard-of-hearing client had trouble understanding her in a sentence where she used diphthongized long vowels, would she exaggerate the vernacular features in an effort to aid comprehension, or would she subconsciously suppress them *because* she was making an effort to be easier to understand? Or, would she find a middle ground of some sort (as our findings may suggest she would) where, for example, she would diphthongize less frequently but more radically? Would she behave differently if she were working in France rather than her native Québec?

5. Conclusion

In investigating two separate style-shifting continua--attention-to-speech and clear-speech--in the context of stratified sociolinguistic variables, we found that a small number of variables pattern in a unique way, where increasing casualness results in different outcomes depending on which continuum is being studied. We then found how one such variable, diphthongization in Québec French, patterns with regards to each continuum, using corpus data coded for style and measures of speech clarity. Finally, we found possible evidence for a way to reconcile the potential conflicts between the two continua, where some speakers could realize the vernacular form less frequently, but to a more radical degree in each token. The notion of multiple, competing speech continua brings with it many interesting possibilities for further experimentation, and we hope to see some of these ideas carried out in the future.

Works Cited

Bickerton, D. (1980). What happens when we switch? *York Papers in Linguistics* 9, 41-56.

Boersma, P. and D. Weenink (2014). Praat: doing phonetics by computer. Version 5.4.08. http://www.praat.org/

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication* 15, 39-54.

Clopper, C. (2009). Computational Methods for Normalizing Acoustic Vowel Data for Talker Differences. *Language and Linguistics Compass* 3/6, 1430-1442.

Côté, M.-H. (2005). *Phonologie française*. Manuscript, Département de langues, linguistique et traduction, Université Laval, Québec, Québec.

Coupland, N. (1984). Accommodation at work: Some phonological data and their implications. *International Journal of the Sociology of Language* 46.

Diehl, R., et al. (1996). On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics*, 24 (2), 187-208.

Douglas-Cowie, E. (1978). Linguistic code-switching in a Northern Irish village: Social interaction and social ambition. In P. Trudgill (ed.), *Sociolinguistic patterns in British English*. London: Edward Arnold, 37-51.

Dumas, D. (1974). Durée vocalique et diphtongaison en français québécois. In: Y.-C. Morin, M. Picard, P. Pupier, and L. Santerre (eds.), Le français de la région de Montréal. Montréal: *Les Presses de l'Université du Québec*, 13–55. In Gess 2008.

Fischer, J. (1958). Social influences on the choice of a linguistic variant. Retrieved from http://web.stanford.edu/~eckert/PDF/fischer1958.pdf/

Gahl, S., et al. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 2012, doi:10.1016/j.jml.2011.11.006.

Gess, R. (2008). More on (distinctive!) vowel length in historical French. *Journal of French Language Studies,* 18, 175-187.

Goldstein, U. (1980). An articulatory model for the vocal tracts of growing children. Doctoral dissertation, MIT. In Diehl et al. 1996.

Hubbell, A. (1950). *The Pronunciation of English in New York City: Consonants and Vowels*. New York: King's Crown Press. In Labov 1972.

Kendall, T. and E. Thomas (2015). Package 'vowels.' Version 1.2-1. http://cran.r-project.org/web/packages/vowels/vowels.pdf/

Kroch, A. (1978). Towards a theory of social dialect variation. *Language in Society* 7, 17-36.

Labov, W. (1972). *Socioliguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H Theory. In Hardcastle, WJ; Marchal, A, eds. (1990). *Speech production and speech modeling.* Amsterdam: Kluwer Academic, 403-39.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America* 49, 606-608.

Mackenzie, L. and Sankoff, G. (2010). A quantitative analysis of diphthongization in Montreal French. *University of Pennsylvania Working Papers in Linguistics*, 5(2): article 11.

Maurais, J. (2008). Les Québécois et la norme: L'évaluation par les Québécois de leurs usages linguistiques. Office québécois de la langue française. https://www.oqlf.gouv.qc.ca/etudes/etude_07.pdf/.

Milne, P. (2013). The variable pronunciations of word-final consonant clusters in a force-aligned corpus of spoken French. PhD thesis, University of Ottawa.

Moon S-J and Lindblom, B. (1989).Formant Undershoot in Clear and citation-Form Speech: A Second Progress Report. *STL-QPSR* 1/1989. 121-123.

New, B. and Pallier, C. (2001). Lexique. 3.80. University of Savoy, Chambéry, France.

Ostiguy, L. and Tousignant, C. (2008). *Les prononciations du français québécois.* Montréal: Guérin universitaire.

Russell, J. (1982). Networks and sociolinguistic variation in an African urban setting. In S. Romaine (ed.), *Sociolinguistic variation in speech communities.* London: Edward Arnold. 125-40.

Santerre, L. and Millo, J. (1978). Diphthongization in Montreal French. In D. Sankoff (ed.), *Linguistic Variation: Models and Methods*, 173–184. New York: Academic Press.

Simpson, A. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass* 3(2), 621-640.

Smiljanić, R. and Bradlow, A. (2009). Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes. *Language and Linguistics Compass* 3(1): 236–264.

Stevens, S. S. and Volkmann, J. (1940). The relation of pitch to frequency: A revised scale. *American Journal of Psychology* 53, 329-353.

Thomas, E. (2010). *Sociophonetics: An Introduction*. Basingstoke: Palgrave Macmillan.

Traunmuller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America* 88, 97-100.

Traunmuller, H. (1997). Auditory scales of frequency representation. http://www.ling.su.se/staff/hartmut/bark.htm/

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society* I, 179-195.

Walker, D. (1984). *The Pronunciation of Canadian French.* Ottawa: University of Ottawa Press.

Wardhaugh, R. (2002). *An Introduction to Sociolinguistics.* Hoboken: Blackwell.

Zwicky, A. (1972). On Casual Speech. *Chicago Linguistic Society* 8, 1972.

Appendix 1. Whitelist of historically-long words found in ANQ data

arrête
empêche
enquêtait
enquête
enquêtes
entêtent
êtes
être
êtres
extrême
extrêmes
fête
mêle
même
mêmes
pêche
prêtait
prête
tête