STUDII DE LINGVISTICĂ

Vol. 11, 2021

Variation phonétique en parole continue
Phonetic variation in continuous speech

Număr coordonat de
Yaru Wu, Martine Adda-Decker,
Lori Lamel, Ioana Vasilescu

Editura Universității din Oradea

Studii de lingvistică 11, 2021, 55 - 84

Hypoarticulation as a tool for assessing social distance: an acoustic study of speech addressed to different types of interlocutors

Mélanie Lancien[1]
Marie-Hélène Côté[2]

Abstract: Work within Hyper-Hypoarticulation Theory (H&H) and Communication Accommodation Theory (CAT) is increasingly focused on the adaptation of speech to the identity of the interlocutor (Koppen *et al.* 2017, Pardo *et al.* 2012, among others). These studies show a correlation between changes in the rate and spectral characteristics of speech (especially vowels) and the relationship between the speakers. Using the Diapix task (Baker & Hazan 2011), 10 Québec-French-speaking couples were invited to interact together and with two strangers, one French and one Québécois. This produced a corpus of 25h of speech and 121000 vowels. Spectral variations (especially hyper- / hypo- articulation), and changes in speech rate depending on the interlocutor, were studied using ((G)LMM) analysis. Our results reveal a correlation between the degree of social distance and speech reduction: the closer the interlocutors are (partners), the more speech is reduced.

Key words: sociophonetics, phonostylistics, Quebec French, vowel variation, acoustics, social distance.

1. Introduction

Phonostylistics and sociophonetics specifically aim at studying phonetic modifications in speech caused by several parameters of the communication situation. In the past, researchers have mostly focused on hypoarticulation (the H&H theory – see Lindblom 1990 – stating that speakers tend to hypoarticulate or hyperarticulate, i.e. reduce or enhance the articulatory effort, according to the situation); see for instance Harmegnies & Poch-Olivé

[1] Université de Lausanne, Section des Sciences du Langage et de l'Information; melanie.lancien@unil.ch.
[2] Université de Lausanne, Section des Sciences du Langage et de l'Information; marie-helene.cote@unil.ch.

1992, 1994; Scarborough & Zellou 2013, Scarborough *et al.* 2007; Simon *et al.* 2010; Grossman *et al.* 2018. From the point of view of acoustic phonetics, research has often been limited to a comparison between the characteristics of read speech and spontaneous speech, or familiar style and formal style. These studies have also often been carried out on very small corpora (such as case studies), sometimes recorded under quite artificial conditions (laboratory conditions, simulated tasks). For now, they indicate that speech tends to be phonetically reduced in informal situations and spontaneous speech (as opposed to formal situations and read speech). This reduction is instantiated by several parameters: vowel formants and duration are modified (e.g. Adda-Decker & Lamel 1999, Rouas *et al.* 2010, Koppen *et al.* 2017), vowels are hypoarticulated or even missing (e.g. Scarborough & Zellou 2013, Ernestus *et al.* 2015), acoustic spaces are smaller (Harmegnies & Poch-Olivé 1992, 1994; Poch-Olivé & Harmegnies 1992; among others), speech rate is faster (e.g. Goldman *et al.* 2010, Grosman *et al.* 2018).

In this paper, we further explore the effect of social distance between speakers on phonetic variation in spontaneous speech. For our specific purpose, social distance is defined following Kelley *et al.* (1983) and work in psychosociology by Becker & Useem (1942), notably. Kelley and colleagues argue that a close relationship between two individuals is characterized by the joint realization of a large number of activities. The closeness of the relationship increases when: 1) the individuals frequently impact each other; 2) the degree of impact for each joint activity is strong; 3) this impact involves various types of activities for each person; and 4) these properties characterize the set of joint activities over a relatively long period of time (Kelley *et al.* 1983: 13). To this definition we add the idea that closeness can vary according to the cultural/social identity of the speakers (developed by Becker & Useem 1942 or Maisonneuve 1966, for instance). We therefore seek to observe how a speaker changes his/her pronunciation according to the kind of relationship s/he has with the partner with whom the speech activity is performed. Our initial hypotheses are the following: each speaker adapts his/her pronunciation to his/her interlocutor and, in terms of hyper-hypoarticulation, this accommodation goes in two opposite directions according to social distance. If we follow the communication accommodation theory (Giles & Smith 1979), people in close relationships (minimal social distance), such as couples (e.g. Becker & Useem 1942, Maisonneuve 1966, Kelley *et al.* 1983), have had time to create particular forms of pronunciation, particular habits, a bit like a jargon but on a phonetic and articulatory level. These articulation habits (here hypoarticulation) are the result of an accommodation process that can be considered as convergent and reciprocal (both adapt to the other). On the other hand, when the social distance is maximal (strangers or even foreigners; e.g. Becker & Useem 1942, Maisonneuve 1966, Kelley *et al.* 1983), accommodation to the other would consist in hyperarticulation, to facilitate understanding. The greater the social distance, the more accommodation (on the acoustic and articulatory level) will consist in a form of hyperarticulation; conversely, the smaller the social distance, the more accommodation will manifest itself in a form of hypoarticulation. Thus, the degree of hyperarticulation would be indicative of the social distance between the two interactors.

2. Theoretical framework

In this work we would like to address the issue of hyper-hypoarticulation and its link to social distance through the prism of two frameworks: the Hyper-Hypoarticulation theory (H&H) and the communication accommodation theory (CAT). Both of these frameworks understand the phonetic modification of speech as being closely related to the needs of the interlocutors and provide complementary points of view from the acoustic and sociopsychological perspectives.

2.1. The Hyper-Hypoarticulation theory (H&H theory)

Lindblom's (1990) so-called H&H model states that clarity and precision of articulation (as measured by the reduction of spectral properties of phones) vary according to two elements: the first one, paralinguistic, is the needs of the interactors for the successful realization of the interaction; the second one, linguistic, concerns speech rate. This model therefore introduces the idea (relatively innovative for phonetics at that time) that speakers of a language adapt the clarity of their speech to the information required by their interlocutor to understand the message. A speaker will thus hyperarticulate, i.e. extend the acoustic space between vowel categories (making them more distinct from each other), when the interlocutor needs a maximum of acoustic information (for instance under noisy conditions), and reduce his/her articulatory effort when the interlocutor is able to fill the (acoustic) information gap with the information s/he already possesses (generally given by the context of interaction).

Different metrics have been proposed to evaluate the degree of hyper-hypoarticulation at a spectral level. The best known one is undoubtedly the size of the acoustic vowel space in a F1xF2 space[3].

[3] A bidimensional space representing the vowel triangle, with F1 values (aperture) on the y axis and F2 values (front-back position of the tongue) on the x axis.

Within this framework, several Euclidean distances can be measured, including the distance between the calculated centers of the vowel categories (themselves calculated by the Euclidean distance between the different phones pertaining to this category) (Gendrot & Adda-Decker 2007), but also the distance of the phones to the calculated center of the system (Harmegnies & Poch-Olivé 1994), or the distance of the phones to the calculated center of their vowel category (Audibert *et al.* 2015).

Experiments in phonostylistics cast in the H&H theory have already demonstrated that speech style, or the conditions in which speech is produced[4], affects the degree of speech hypoarticulation. Most of these studies have dealt with spontaneous *vs* read speech and have shown that spontaneous speech is more hypoarticulated than read speech (Harmegnies & Poch-Olivé 1992, Adda-Decker & Lamel 1999, Rouas *et al.* 2010). More recent work also suggests that beyond this dichotomy, there seems to be a gradation, a continuum of "preparation" or "casualness" that correlates with the hypo-hyper continuum and the speech rate continuum (Harmegnies & Poch-Olivé 1994, Hupin & Simon 2007, among others). The more speech is produced in natural, ecological, spontaneous conditions, the more vowels are hypoarticulated and the rate is fast. Besides, studies such as Koppen *et al.* (2017) have pointed toward the fact that hypoarticulation and faster pace could be linked to the social distance between speakers. Their work on speech addressed to friends *vs* experimenters, as well as studies by Yuan *et al.* (2006) or Scarborough & Zellou (2013) on speech addressed to different kinds of interlocutors, tend to demonstrate that the closer one is to his/her fellow speaker (in terms of social relationship, cultural origins, and mother tongue) the faster s/he speaks and the more s/he hypoarticulates. Such results also arise from work casted in the CAT framework, as discussed below.

2.2. Communication accommodation theory

Psychosociology and communication accommodation theory (Giles & Smith 1979, among others) also provide an interesting basis for the exploration of the effect of social distance on speech. CAT researchers define the accommodation process as: "the way interactants adjust their communication behaviors to [...] their fellow speakers" (Gasiorek *et al.* 2015: 2). Among the four fundamental principles of CAT, as defined by Giles & Ogay (2007), three are specifically relevant to our work. 1) The social affiliations of individuals are often negotiated during interactions. 2) Speakers

[4] The definition of "style" being quite fuzzy and variable across authors, we choose to stick with the notion of "condition(s) of production" rather than "style" or "phonostyle".

have expectations regarding optimal levels of accommodation (based on stereotypes about the members of different groups and social and situational norms). 3) Speakers use specific communication strategies (in particular convergence and divergence) to signal their attitudes toward their fellow speaker and their respective social groups. Social interaction is thus a subtle balance between the needs for social inclusion and differentiation (assertion of one's own identity).

Experiments carried out in psychosociology and the CAT have largely focused on dialectal accommodation and have given more consideration to perception than to production. However, studies such as Pardo *et al.* (2006, 2018), Yuan *et al.* (2006), Delvaux & Soquet (2007) or Aubanel & Nguyen (2020) have showed that speakers adapt their F0 (pitch of the voice), intensity, and speech rate to those of their fellow speakers. For instance, Pardo *et al.* (2012), in a study of the speech produced by five dyads of roommates, observed that over the course of an academic year, roommates show signs of both convergence and divergence in the pronunciation of certain words during reading tasks. The results of an acoustic analysis show that the Euclidean distance between the vowels produced by each pair of roommates appears to be shorter between the first and second recording sessions, then increases between the second and third sessions (presumably due to vacation) and decreases again between the third and fourth sessions. In the continuation of this work, Pardo *et al.* (2017) measured the distance between the durations, F0, F1, F2, and the vowel space (on a F1xF2 space) of the productions of 108 speakers. The study reveals that convergence, in its acoustic dimension, is mainly expressed by vowel duration, the results on the F1xF2 space and F2 being marginal, and no significant results being found for F0 and F1. However, the same study showed that all these variables appear as predictors of the results of AXB perception tests[5] of the degree of convergence. Other works on this topic, such as Aubanel & Nguyen (2020) or Delvaux & Soquet (2007), concerning a more immediate accommodation, and generally in less ecological experimental settings, have also highlighted an unconscious phenomenon of instantaneous convergence of a speaker's speech toward that of the people around him/her. Delvaux & Soquet (2007) set up a shadowing task whose results showed a convergence in the temporal (vowel duration) and spectral (F1, F2, F3, and MFCCs[6])

[5] Tests where the listener has to decide whether word X (the target) sounds more like word A or word B.

[6] Mel frequency cepstrum coefficients (MFCCs) are coefficients that represent the short-term power spectrum of a sound, and thus offer a representation of that sound's properties.

measurements of vowels produced during a word repetition task. Although limited in terms of the phonetic material observed, this study is particularly interesting since it compares word repetitions in two situations: one where the participant is in the presence of an experimenter and one where s/he hears speech through loudspeakers. She has to repeat words previously pronounced by the experimenter or loudspeakers with a different regional accent from that of the participant. The results show a convergence towards (or, according to the authors, an imitation of) the speech produced by both the real and audio persons. The authors also note that the main index of convergence is vowel duration. In the same vein, more recent work by Remysen (2020) showed an accommodation from Quebec French speakers to France French speakers that could be interpreted as accommodation in the degree of hypoarticulation (most likely driven by linguistic insecurity). In this work, the author found that Quebec French speakers were less likely to reduce consonant clusters when addressing a speaker from France than one from Quebec.

Finally, a series of studies conducted in the 1990s in the United States by Gregory and colleagues using experimental protocols derived from sociolinguistic methods (interviews) show that physical (as opposed to virtual) presence leads to greater convergence (in F0, intensity or rhythm) (Gregory & Hoyt 1982, Gregory 1990, Gregory *et al.* 1997, 2001). These results are congruent with more phonetically-oriented work by Scarborough *et al.* (2007) and Scarborough & Zellou (2013) on adaptation (or accommodation) to the interlocutor, showing an effect of task authenticity. Lastly, Gregory & Webster (1996) also show that in an interaction, speakers in a lower position accommodate (in terms of low frequency range) to a dominant partner (here a celebrity). The results of this study reinforce the legitimacy of the CAT principles by providing an experimental demonstration and enlighten the need to proceed to both ecological and non-ecological studies to better understand these phenomena.

3. Experiments and methods

To investigate the effect of social distance and the interlocutor's identity on speech reduction, we propose to cut the social distance scale (originally continuous) into discrete categories that will be represented by one interlocutor each (very close, stranger from the same regional community, stranger from a different regional community). We analyze acoustic features (vowel duration, vowel dispersion in the acoustic space, and compacity) representing the degree of hyper-hypoarticulation in the speech of the speakers interacting with the three interlocutors. Our corpus includes 20 speakers of Quebec French (recorded by the 1st author at the University of Quebec at Chicoutimi – UQAC – in 2019), placed in four situations: 1) reading, 2) identifying differences in images (Diapix task, Baker & Hazan 2011) with a spouse, 3) identifying those differences with a stranger from the same regional origin (male, 26 y.o.), and 4) identifying those differences with a stranger from France (female, 25 y.o., with a "standard"[7] accent). The reading task is a control situation, the other three correspond to steps on the social distance scale. The following sections describe our experimental framework, speakers' profiles, and our methodological choices regarding vowel categorization, the metrics used to analyze the speech material, and data processing.

3.1. Data collection

In order to study the correlations between speech reduction and social distance, we need to collect speech as natural as possible while making different pairs of speakers interact over the same material. The tasks proposed by the Diapix protocol (Baker & Hazan 2011) fulfill these conditions; they were complemented by the reading of 482 key words as a control task[8]. In the interactive tasks, presented as a game, two speakers have identical pictures except for 12 differences which they have to identify without showing each other's image. An example of those pictures is given in Figure 1. The images, which show scenes at the farm, the beach and in the city, have been carefully designed to induce specific lexical items (target words), for example the words "sheep", "tree" and "tractor" for images representing a farm. This game has already been used as an experimental protocol with participants of all ages (ranging from 8 to 85 years old). In addition, work on convergence cited above, such as Pardo *et al.* (2018, 2019), have already reported the efficiency of such types of tasks and their practicality from a methodological point of view for the study of accommodation to the interlocutor. Thanks to this task, the frame and purpose of the interactions remain constant: no spectators, same room, same material, same temporal organization, same activities, same goals, same instructions (see Brown & Fraser's (1979) taxonomy of a communication situation). Only the participants and their relationship changed between the different stages of the protocol. A total of 12 pairs of images were used by all participants, four

[7] By "standard" we mean corresponding to the European norm in the representation of Québécois speakers.

[8] During the reading task speakers were alone and were not listened to (which they were aware of).

pairs in each of the scenes (beach, farm, street). In order to be used with French speakers in Quebec, the images were adapted, mainly by translating the texts present on the images from English into colloquial Quebec French.

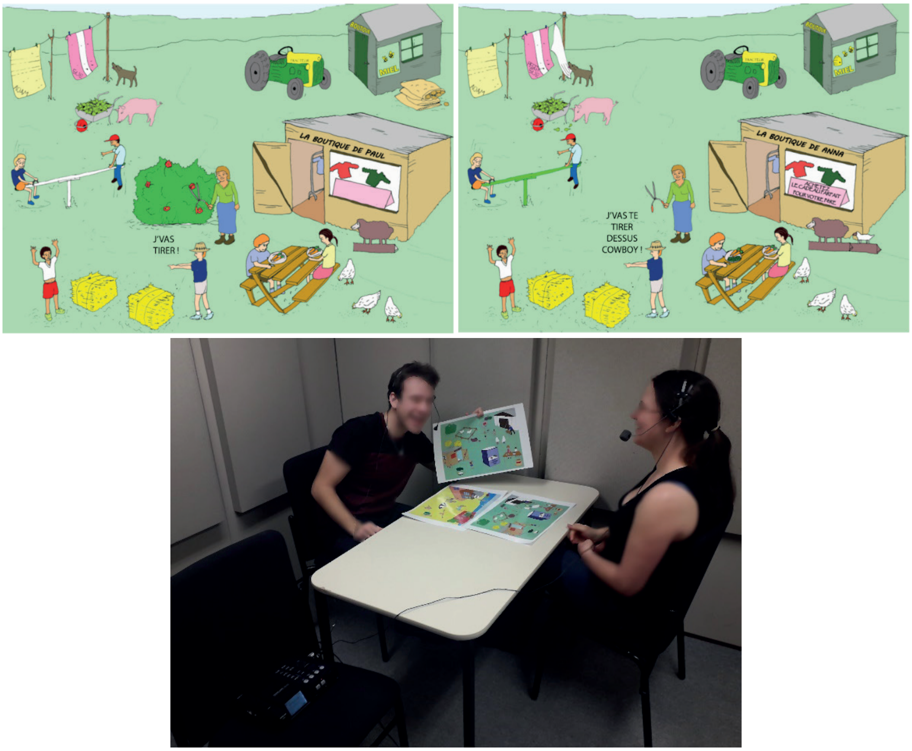

Figure 1: Example of pictures from Diapix given to pairs of speakers (above) and their use during a recording session (below).

One recording session lasted about 3h, during which our participants played with their spouse, with an unknown experimenter from Saguenay-Lac-Saint-Jean (Quebec, Canada), and with an experimenter from France who had never met the speakers before either. At the end they read a list of words and sentences based on the target words elicited by the Diapix pictures[9]. Overall, 5.91h of speech with a spouse (*Couple* hereafter), 6.35h of speech with the local stranger (*Local*), 6.28h of speech with the foreigner (*French*) and 5.66h of read speech (*Reading*) were recorded.

[9] Note that these data are part of a larger study also involving a condition where speech is uttered alone (see Lancien 2021). More results, mainly regarding rhythm, are also available in this thesis.

3.2.Speakers' profiles

Our 20 speakers (10 heterosexual couples, so 10 women, 10 men) were all Quebec French speakers from the area of Saguenay-Lac-Saint-Jean (hereafter SLSJ). The average age was 29.35 years old (min =20, max = 51, median = 28, σ[10] = 7.28). All couples had been together for more than a year at the time of recording and the majority of them had been together for more than five years. The participants' level of education and socio-economic characteristics are much less homogeneous, since our pool includes students as well as doctors, employees or civil servants. Their geographic origin is however very homogeneous, since 14 of them came from the Saguenay agglomeration (146 920 inhabitants in 2019[11]) and 17 had spent the majority of their lives in the SLSJ region. With the exception of those with university education in other regions of Quebec, they mostly remained in the region of origin of their parents (SLSJ). Therefore, age (too homogeneous) and social class (too heterogeneous) could not be included as social variables in the following analysis.

3.3.Vowel categories

Quebec French has a very rich vowel system, with a larger phoneme inventory than in Standard French and distinct allophones for most phonemes, depending on syllable structure and position in the word (e.g. Walker 1984, Côté 2012). We adopt the phoneme inventory in Table 1, comprising 16 oral vowels and four nasal vowels (which will be ignored here). This set excludes schwa, which is realized as [œ] in Quebec French (Séguin 2010). The mid and low vowels contrast in length as well as quality: high-mid vowels, /ɜ/ and /ɑ/ are long, /ɛ, œ, ɔ, a/ are short. However, contrasts among mid and low vowels are neutralized before final /ʁ/ and word-finally. Before /ʁ/, only /ɛ, œ, ɔ, ɑ/ appear and vowels are lengthened and diphthongized (e.g. *part* 'part' /pɑʁ/ [pɒ͡oʁ]); word-finally we find essentially /e, ø, o, ɑ/.

	Front unrounded	Front rounded	Back
High	i	y	u
High-mid	e	ø	o
Low-mid	ɛ, ɜ, ɛ̃	œ, œ̃	ɔ, ɔ̃
Low	a		ɑ, ɑ̃

Table 1: Phoneme inventory of Quebec French (cf. Walker 1984)

[10] Sigma represents the standard deviation from the mean.
[11] https://statistique.quebec.ca/fr/produit/tableau/estimations-de-la-population-des-mrc.

To understand the allophonic realizations of these vowels, it is useful to divide the oral vowels into three subsets: the high vowels, the short mid and low vowels, and the long mid and low vowels. Four contexts are relevant: in open syllables, in syllables closed by /ʁ/, in syllables closed by /v, z, ʒ/ (voiced fricatives), and in syllables closed by other consonants. The generalizations are as follows:

- Long mid-low vowels diphthongize in all closed syllables;
- Other vowels lengthen before /ʁ, v, z, ʒ/ and diphthongize before /ʁ/ (e.g. *fête* 'party' /fɜt/ [fa͡ɛt] vs. *faites* 'you.PL do' /fɛt/ [fɛt]);
- High vowels lax to [ɪ, ʏ, ʊ] before consonants other than /ʁ, v, z, ʒ/. They also undergo regular devoicing in non-final syllables in the context of voiceless consonants.

These processes give rise to four series of allophones for high vowels (disregarding devoicing): diphthongizing before /ʁ/, long and tense before /v, z, ʒ/, lax before other consonants, and short and tense in open syllables. Short mid and low vowels have three of these four allophonic categories (excluding lax variants). Long mid and low vowels have two major allophones: diphthongizing in closed syllables and not diphthongizing in open syllables. In addition, /ɑ/ has a specific word-final allophone close to [ɔ]. In terms of their surface realizations, the 12 oral vowels are therefore associated with more than 30 vowel categories with distinct qualities; these are referred to as vowel classes. These classes are defined by the segmental and prosodic context, but their realization is also partly influenced by social factors (e.g. Paradis 1985). Some serve as sociolinguistic variables, notably the diphthongized variants, the word-final allophone of /ɑ/, and the lax high vowels (e.g. Chalier 2019, 2021). Thus, Quebec French (QF) offers a rich material to study sociophonetic variation. Moreover, this variety has been little studied in phonostylistics and the complexity of its system raises new methodological questions, notably in the treatment of diphthongizing vowels in French.

Given their variability of realization, phonemes do not appear to be the right unit of analysis for a study focusing on vowel duration and quality. Thus, we use vowel classes, but the classes defined above could not all be retained, due to the low frequency of some of them. A few classes were merged when they involved reasonably close vowel types, other classes were simply removed. Word position could not be taken into account, for most classes did not show up in every position for every speaker in every condition. So, no distinction is made between final and non-final syllables. We ended up with 16 vowel classes, each containing at least four occurrences from each participant in each condition. These classes are as follows: i#R, iK, y#R, yK, u#R, e#, ɛ#, ɛR, ɛK, ø#, o#. ɔ#, ɔK, a#, aC, ɑ#, where # indicates a syllable boundary (anywhere in the word), R refers to the set of lengthening consonants /ʁ, v, z, ʒ/, K to the set of non-lengthening consonants, and C to any consonant. The classes i#R, y#R and u#R include high vowels in open syllables as well as syllables closed by a lengthening consonant /ʁ, v, z, ʒ/[12].

3.4. Data processing and final data set

The collected 25h of audio data were first segmented into interpausal units (IPUs) – defined as chunks of continuous speech – and silent pauses – defined as silences longer than 200ms – using the SPPAS software (Bigi 2012, Lancien *et al.* 2020). This aligned segmentation was then corrected by hand and the IPUs were subjected to a hand-made augmented orthographic transcription following the standard spelling rules, with liaisons, phone elisions and intra-IPU pauses noted according to the conventions provided by SPPAS. SPPAS was then used to perform an aligned segmentation of the signal into phones, syllables, and words, as well as a series of time analyses (such as computing the number of syllables/IPU).

To extract the acoustic parameters of vowels, a Praat script[13] was used. It allowed the extraction of vowels' mean F1 and F2 (averaged over the vowel length) and mean duration, as well as the label of the vowel, its left and right context, the label of the word, the content of the IPU, the duration of the IPU, and the number of syllables in the IPU. The dataframe generated by this script was then filtered to remove the main formant detection errors and alignment errors (visible in the duration of the vowels). This filtering was made by a R script we wrote to compute Mahalanobis[14] distances between phones and the mean distribution of the vowels' duration, F1 and F2. This measure calculates the distance between a point P and the mean of a distribution D; in our case P is a vowel exemplar and D the distribution of variable X for the group to which P belongs (for example, P is an occurrence of [i] and D is the distribution of F2 for class iK in the reading condition). The Mahalanobis distance provides a multidimensional measure of the number of standard deviations between P and the mean of D. This distance increases as P deviates from the average in a given space, in our case a three-

[12] From now on we use SAMPA to refer to these classes.

[13] © Nicolas Audibert, slightly adapted to our needs.

[14] The Mahalanobis distance from a multivariate vector x = (x1,x2,x3, . . . ,xp)T to a set of vectors of mean values μ = (μ1,μ2, μ3, ... ,μp)T and having a covariance matrix ∑ is defined as follows: Dx = (x - μ)∑ -1(x - μ). We computed it thanks to the mahalanobis () function of stats package (v3.6.3) from R (v3.6). The square root of the result gives a number of standard deviations.

dimensional space consisting of the duration of the phone (s), its mean F1 (Hz), and its mean F2 (Hz). More precisely, we felt that the best filtering was the one for which μD was calculated according to vowel class and condition of production. With this grouping, and choosing the threshold of d≤3 (i.e. 1.7 standard deviation from the mean) to separate "valid" data points from noise, the vast majority of unrealistic values of duration, F1 and F2 were removed from our data, and we had enough vowel exemplars per class to meet the needs of the calculations and analyses described in the following sections (min. 3 phones per class, per speaker and per condition). This method is not perfect and still fails to exclude a few extreme values, but it seemed to us to be a valid alternative (or complement) to the filters based on fixed ranges of formant values (such as the one used by Gendrot & Adda-Decker 2005).

After the filtering process, our dataset consisted of 109 134 vowel occurrences from the 16 classes listed in 3.3. On average, vowel classes include 6 820.8 occurrences (with σ= 5 918, min = 2 099, max = 21 909), all speakers and all conditions pooled. The average number of vowel tokens in each condition, for all classes and all speakers, is 27 283.5 (with σ=1 644.4, min = 24 594, max = 29 046). Finally, each speaker produced between 4 and 538 vowel occurrences per class in each production condition, with an average of 5 456.7 occurrences produced by each speaker (σ = 1 030.9, min = 3 825. max = 7 612), all conditions and all classes pooled. Figure 2 illustrates the percentage of phones realized in each condition for each vowel class.

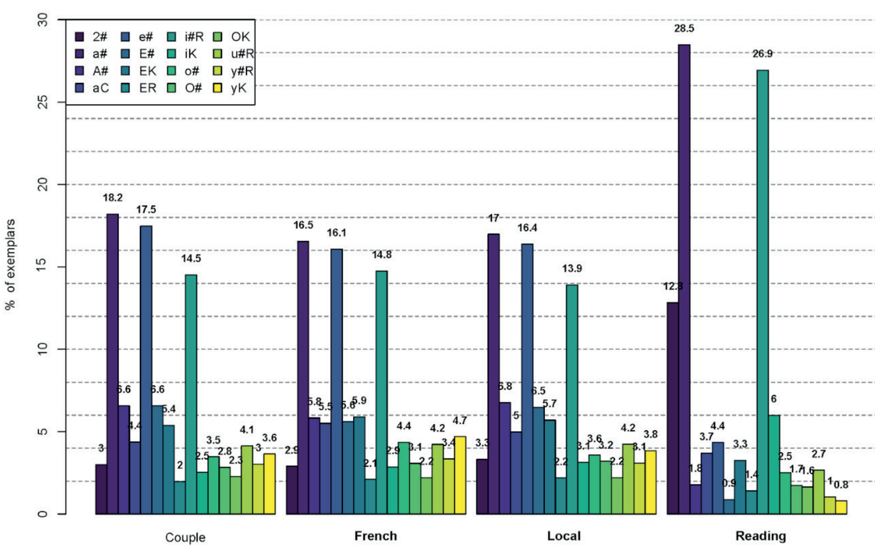

Figure 2: Percentage of phones from each vowel class produced in each speaking condition (all speakers pooled)

3.5. Selected acoustic features

We propose to study speech reduction through variations in vowels' duration, distance to the centroid of each speaker's vowel system and the centroid of their class, F1-F2 compacity, and speech rate (in syll/sec). We chose to transform frequencies using Lobanov's (1971) method, as Lobanov's *z*-scores have been shown to be the best normalization unit to preserve the identity of the vowel category as well as sociophonetic information, while reducing physiological variations (Adank *et al.* 2004, Fabricius *et al.* 2009, Flynn & Foulkes 2011). Thus, formant frequencies in Hz were transformed into *z*-scores before being used to compute the distances and compacity metrics.

Vowel duration was extracted by the previously mentioned Praat script; the two distances and compactness were computed on the basis of the mean F1 and F2 values extracted by the same script and transformed into *z*-scores. Regarding distances to the computed center of the system and the computed center of the class, we followed Harmegnies & Poch-Olivé (1992) and Huet & Harmegnies (2000) and their index of centralization of the system based on a measure of the deviation of a formant value (F1 or F2) from a reference mean and on a measure of Euclidean distance between a vowel (or the center of gravity of a vowel category) in the F1xF2 space and a reference center of gravity. This measure allows us to quantify the inter- and intra-category variation. The centroid of the vowel system of each speaker is calculated on the basis of the distances between different phones or categories of phones, then the degree of centralization of a phone or category is estimated by the distance between this phone or category and the calculated center. Here we consider the distance of the phones to the centroid of the system for each speaker, but also the distance of the phones to the centroid of their vowel class. These distance measurements allow us to evaluate the degree of centralization (or reorganization) of the vowel system, a large dispersion with respect to the centroid of the system being indicative of increased hyperarticulation (more distinctiveness between phones), and a greater distance to the centroid of the class being indicative of increased hypoarticulation (less homogeneous categories). As for the measure of compacity, it simply corresponds to the difference between the extracted mean F2 and F1 values (C = F2-F1). This measure aims at qualifying and quantifying the posteriorization or anteriorization movements of vowels from classes such as AC (/ɑ/ within a syllable closed by any consonant), which are known to fluctuate. Lastly, speech rate was calculated in syllables per second for each IPU. The number of syllables per IPU and the duration of IPUs were computed by the TGA (Time Group Analysis) tool from SPPAS.

4. Results and analysis

To analyze vowel features and speech rate, we computed mixed models (LMM or GLMM, depending on the diagnostics and distributions of variables – see Bates *et al.* 2015) under R 3.5 (R Core Team 2018), except for vowel duration, for which we used a clustering method (Lê *et al.* 2008). The metrics were set as dependent variables (DV). For the vowel spectral measures, CONDITION, SEX, VOWELCLASS and their interactions were set as fixed independent variables (IV), and SPEAKER and WORD as random IVs. Given the importance of accentuation and syllable position (notably with respect to diphthongization and laxing in QF), we tried to set syllable position as another independent variable. The variable was unfortunately defective (for most classes fewer than 3 phones occurred in every syllable position in every production condition for every speaker) and had to be removed. The following sections detail the models and results regarding the effect of the production condition on the spectral and temporal metrics described above. All graphics were realized under R (R Core Team 2018) using the *ggplot2* (Wickham 2016) and *emmeans* (Lenth 2020) packages.

4.1. Vowel duration

For the analysis of vowel duration, we performed a clustering and then tested the difference between the clusters with an ANOVA. Our results show that vowel duration split into three groups ($p < 0.001$) corresponding to 1) speech addressed to a spouse (the shortest), 2) speech addressed to a stranger (both *Local* and *French*), and 3) read speech (the longest). Figure 3 shows the mean vowel duration (in seconds) observed in each condition (all other factors pooled), and Table 2 shows the mean vowel duration in each condition. Vowel length was also dependent on vowel class and speaker, but we won't develop on these factors here.

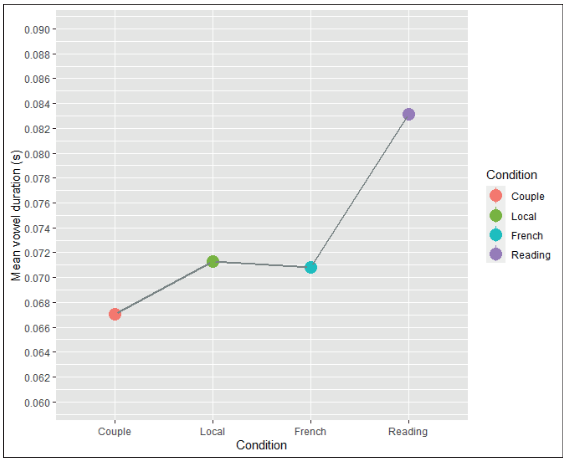

Figure 3: Mean vowel duration (in seconds) observed in each condition, all speakers and classes pooled

Condition	Mean Dur.	Sd Dur.
Couple	0.067	0.036
Local	0.071	0.040
French	0.070	0.039
Reading	0.083	0.036

Table 2: Mean and standard deviation values for vowel duration, for each condition (in seconds)

4.2.Distance to the centroid of the system

The LMM results on the distance to the centroid of the system (on a F1xF2 plane, calculated from the *z*-scores formant values) show a significant main effect of the CONDITION ($x^2(3)=63.5$, $p<0.001$), SEX ($x^2(1)=19.71$, $p<0.05$), CLASS ($x^2(15)=4800.2645$, $p<0.001$), and the interactions between CONDITION and SEX ($x^2(3)=120.54$, $p<0.001$), SEX and CLASS ($x^2(15)=1254.18$, $p<0.001$), and CONDITION and CLASS ($x^2(45)=1599.15$, $p<0.001$). We will not detail here the interactions between CONDITION and CLASS, and we leave aside the interactions not involving CONDITION. Regarding the effect of the CONDITION alone, post-hoc tests (Tukey HSD – computed thanks to the *emmeans* package; see Lenth 2020) show a significant difference between all the conditions ($p<0.005$). As illustrated in Figure 4, the distance to the centroid of the system was the smallest when speech was addressed to a spouse, greater when speech was addressed to the *Local* experimenter, and greater still when speech was addressed to the *French* experimenter. Unsurprisingly, the greatest distance to the centroid of the system was observed in the reading task.

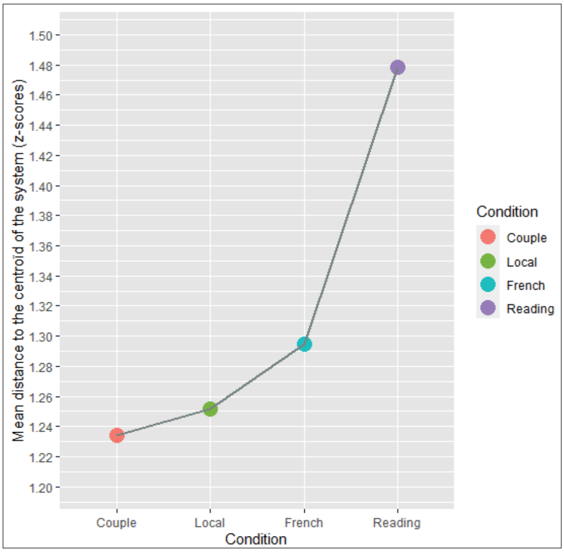

Figure 4: Mean values of the distance of phones to the centroid of the system for each condition of production (all speakers pooled, Euclidean distances based on F1 and F2 in z-scores)

Condition	Mean DistSyst	Sd DistSyst
Couple	1.23	0.569
Local	1.25	0.590
French	1.29	0.604
Reading	1.48	0.611

Table 3: Mean and standard deviation values of the distance of phones to the centroid of the system (Euclidean distances based on F1 and F2 in *z*-scores) for each condition of production

As the interaction between SEX and CONDITION was significant ($x^2(3)=120.5438$, $p<0.001$), we compared the degree of centralization of the system in women's and men's productions. As shown in Figure 5 and Table 4, men and women show a different pattern regarding the two experimenters. First, men centralized more with the *Local* experimenter than women did ($p<0.001$). In men's speech centralization was not greater in the *Couple* condition than in the *Local* condition (actually *Local* was a little more centralized with $p=0.01$). Moreover, the difference between the mean degree of centralization in the *Couple* and *French* conditions was observed to be greater for women then for men (♀: *French* - *Couple* = 1.30 - 1.23 = 0.07; ♂: *French* - *Couple* = 1.29 - 1.24 = 0.05). This difference might not be significant though, considering that ♀*French* was not significantly different from ♂*French*, and ♀*Couple* not significantly different from ♂*Couple*.

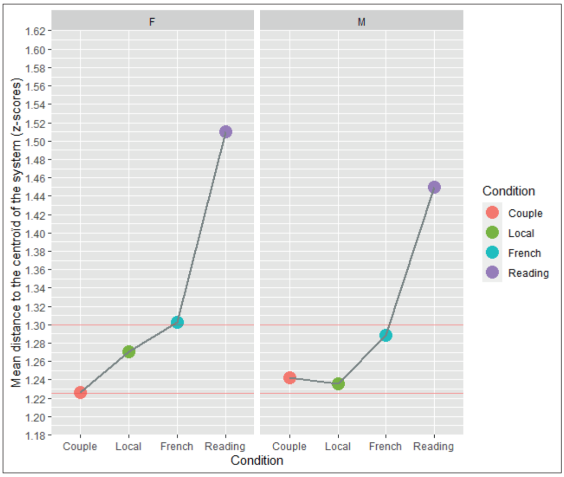

Figure 5: Mean values of the distance of phones to the centroid of the system in each production condition, for male (right) and female speakers (left) (all speakers pooled, Euclidean distances based on F1 and F2 in *z*-scores)

Sex	Condition	Mean DistSyst	Sd DistSyst
F	Couple	1.23	0.539
M	Couple	1.24	0.596
F	Local	1.27	0.561
M	Local	1.24	0.614
F	French	1.30	0.574
M	French	1.29	0.629
F	Reading	1.51	0.579
M	Reading	1.45	0.637

Table 4: Mean and standard deviation values of the distance of phones to the centroid of the system (Euclidean distance based on *z*-scores values of F1 and F2) in each production condition and for each sex

4.3. Distance to the centroid of the vowel class

The GLMM results on the distance to the centroid of the class (on a F1xF2 plane, calculated from the *z*-scores values of formants) show a significant effect of the CONDITION ($x^2(3)=574.4490$, $p<0.001$), CLASS ($x^2(15)=940.95$, $p<0.001$), and the interactions between CONDITION and SEX ($x^2(3)=8.5$, $p<0.05$), SEX and CLASS ($x^2(15)=66.02$, $p<0.001$), CONDITION and CLASS ($x^2(45)=1413.49$, $p< 0.001$), and CONDITION, SEX and CLASS ($x^2(45)=107.42$, $p<0.001$). Once again, we'll leave aside the interactions that don't involve CONDITION, and we won't detail the one between CLASS and CONDITION as well as the triple interaction (too complex to be discussed here).

Regarding the effect of the condition, the post-hoc tests (Tukey HSD) show a significant difference between all the conditions ($p \leq 0.05$) except for the *Couple*-*Local* and *Local*-*French* pairs ($p > 0.1$). The distance to the centroid of the class was the smallest when speech was addressed to a spouse, it was a little greater (not significantly) when speech was addressed to the *Local* experimenter, and significantly greater when speech was addressed to the *French* experimenter then when addressed to a spouse, but not significantly greater than when addressed to the *Local* experimenter. *Reading* showed the greatest distance to the centroid of the class. These results are given in Figure 6 and Table 5.

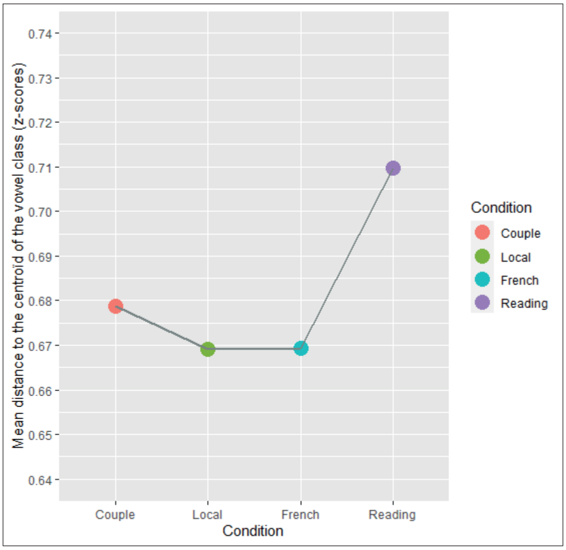

Figure 6: Mean values of the distance of phones to the centroid of the vowel class for each condition of production (all speakers and vowel classes pooled, Euclidean distances based on z-score values of F1 and F2)

Condition	Mean DistClass	Sd DistClass
Couple	0.679	0.416
Local	0.668	0.415
French	0.669	0.415
Reading	0.710	0.451

Table 5: Mean and standard deviation values of the distance of phones to the centroid of the vowel class for each condition of production (Euclidean distances based on *z*-score values of F1 and F2)

Results on the interaction between CONDITION and SEX mainly show differences in the way men and women acted with the experimenters. As illustrated in Figure 7 and Table 6, women showed no difference between *Couple* and *Local*, and little to no difference between *Local* and *French* (p=0.08). For men, *Couple* wasn't different from *French* and *French* wasn't different from *Local*, but there was a greater difference between *Couple* and *Local* (p=0.02). For both sexes the three interactive conditions displayed a smaller distance to the centroid of the vowel class than *Reading*. *Reading* showed no difference linked to speakers' sex, as did *Local*. Men in *Couple* showed little to no difference from women in *Couple* (p=0.07), but for the *French* condition, women displayed a greater reduction of the distance than men (p=0.02).

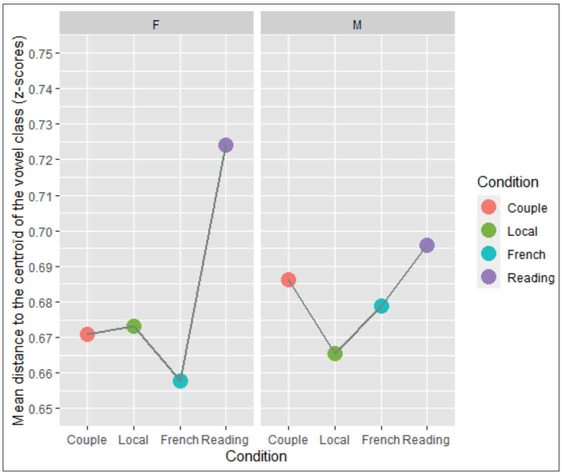

Figure 7: Mean values of the distance of phones to the centroid of the vowel class for male (right) and female speakers (left) (all speakers pooled, Euclidean distances based on z-score F1 and F2 values)

Sex	Condition	Mean DistClass	Sd DistClass
F	Couple	0.671	0.396
M	Couple	0.686	0.435
F	Local	0.673	0.396
M	Local	0.665	0.430
F	French	0.658	0.382
M	French	0.679	0.440
F	Reading	0.724	0.450
M	Reading	0.696	0.451

Table 6: Mean and standard deviation values of the distance of phones to the centroid of the vowel class (Euclidean distances based on z-score F1 and F2 values) in each production condition and for each sex

4.4. Compacity

The LMM run for compacity showed a significant effect of Condition ($x^2(3)=79.22$, $p<0.001$), Sex ($x^2(1)=2.59$, $p<0.001$), and Class ($x^2(15)=36580.94$, $p<0.001$), as well as the interactions between Condition and Sex ($x^2(3)=113.84$, $p<0.001$), Sex and Class ($x^2(15)=449.6$, $p<0.001$), and Condition and Class ($x^2(45)=1319.32$, $p<0.001$). Once again, we'll ignore the interactions that don't involve Condition and won't detail the interaction between Class and Condition and the triple interaction.

As shown in Figure 8 and Table 7, compacity increases as social distance decreases. Post-hoc tests showed a significant difference between every condition except *Couple* and *Local*. Thus, vowels in these two conditions were more compact, which could mean that the system was more posteriorized than in *French* and *Reading*. These results could indicate that speakers adapted their speech to the formality of the task and to the origin of their fellow speaker, thus fronting their vowels to reach qualities closer to those from France (“standard”) French. Very little work has been done on how Quebec French speakers accommodate to Francophones from Europe, but the hypothesis that speakers tend to converge towards the French experimenter's vowel quality is at least consistent with Remysen (2020).

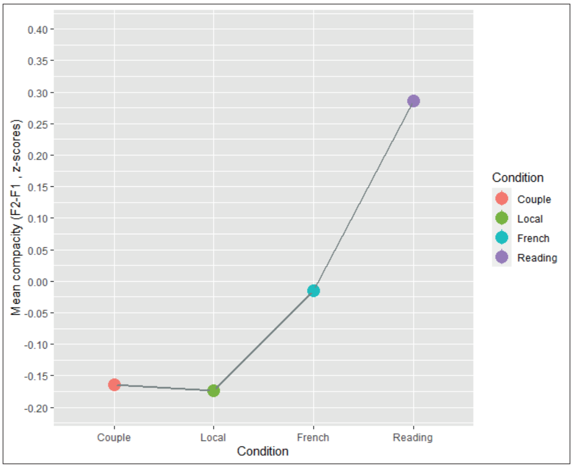

Figure 8: Mean compacity for each condition of production (all speakers and vowel classes pooled, measures based on z-score values of F1 and F2). The greater the mean the less compact the vowels

Condition	Mean Comp	Sd Comp
Couple	-0.164	1.55
Local	-0.174	1.58
French	-0.014	1.63
Reading	0.286	1.85

Table 7: Mean and standard deviation values of compacity in each condition of production (F2-F1 in z-scores)

Here we observe an effect of origin more than social distance (as we defined it). Diffuse vowel classes 2#, a#, e#, E#, EK, i#R, yK, and iK were more compact in speech addressed to the *Local* experimenter than in speech addressed to the *French* experimenter. These results could be interpreted either as a phenomenon of hypoarticulation when addressing the *Local* experimenter (target was not reached), or a fronting movement triggered by a tendency to accommodate to the standard French system (which is undergoing a fronting process, back productions being socially devalued, see Léon & Léon 1997 or Lyche 2010). What was most interesting and very consistent with the hypothesis of fronting was the fact that the intrinsically compact class A# was fronted when addressing the *French* experimenter but not in *Couple* or *Local* (see Figure 9). This could be related to the loss of the /A/ phoneme in most varieties in France, unlike Quebec French, where the /a/-/A/ opposition is very stable (Santerre 1974, 1976, 1979; Martin 1998, 2002)[15].

[15] These observations regarding the French and Local experimenters' systems are based on the information found in the literature, not on the actual observation of these speakers' specific pronunciation.

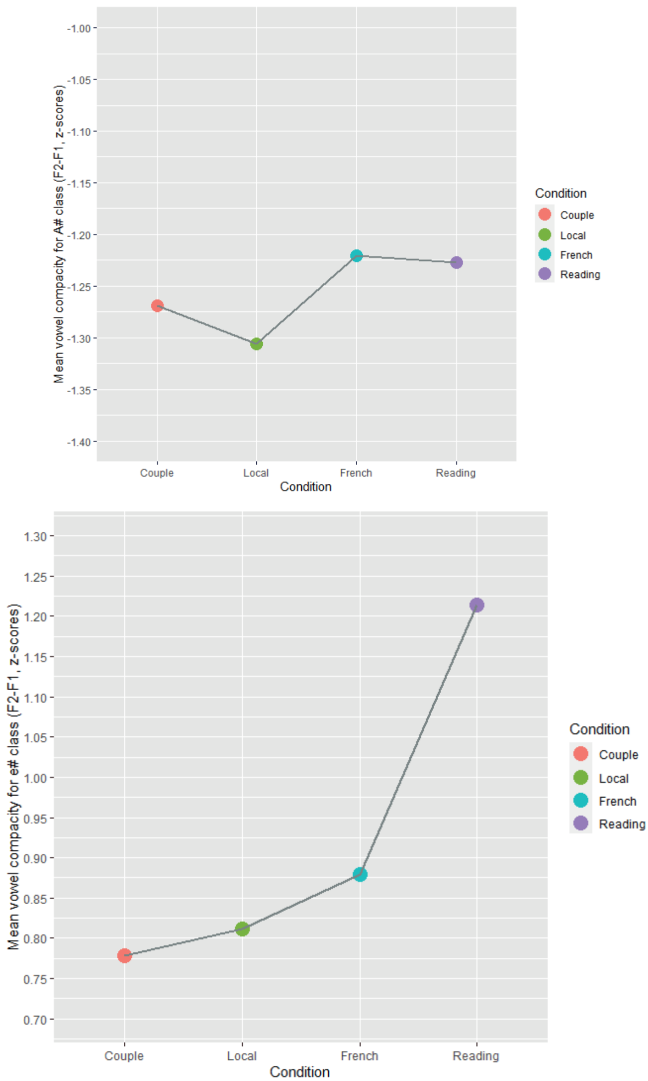

Figure 9: Mean compacity for each condition of production (all speakers pooled, measures based on z-scores values of F1 and F2) for the A# (left) and e# (right) classes. The greater the mean the less compact the vowels

As for the interaction between Sex and Condition, post-hoc tests showed that for both men and women the difference between *Couple* and *Local* was not significant. For both sexes *Couple* was different from *French* (p<0.001), *French* from *Local* (p<0.001), and *Reading* from the three interactive situations (*Couple*, *French* and *Local*; p<0.001). In addition, there were no differences between men and women in each of the conditions: ♂*Couple* = ♀*Couple*, ♂*Local* = ♀*Local*, ♂*French* = ♀*French*, and ♂*Reading* = ♀*Reading*. Figure 10 and Table 8 show the pattern observed for this interaction: *Couple* = *Local* < *French* < *Reading* (from the most compact to the most diffuse).

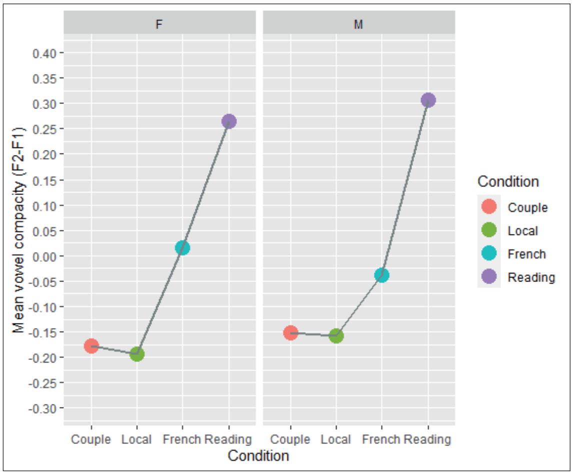

Figure 10: Mean compacity (F2-F1, z-scores) in each production condition and for each sex

Sex	Condition	Mean Comp.	Sd Comp.
F	Couple	-0.177	1.51
M	Couple	-0.152	1.59
F	Local	-0.194	1.57
M	Local	-0.157	1.60
F	French	0.014	1.60
M	French	-0.039	1.65
F	Reading	0.264	1.87
M	Reading	0.307	1.83

Table 8: Mean and standard deviation values of compacity (F2-F1, z-scores) in each production condition and for each sex

4.5.Speech rate

Speech rate was computed in number of syllables/second. Following the observation of data distribution, we chose to exclude measurements lower than 2 syllables/second and higher than 15 syllables/second, so that the model would not be too much influenced by rare and extreme values. For this DV we computed a GLMM with Sex and Condition set as fixed IVs (as well as the interaction between both factors) and Speaker and IPU set as random IVs. The results show a main effect of the production condition on speech rate ($x^2(3)$= 161.37, p<0.001), as well as a main effect of the interaction between Sex and Condition ($x^2(3)$= 12.79, p=0.005). The post-hoc tests performed on the effect of Condition show a significant difference between all production conditions (p ≤ 0.04), leaving us with the following pattern: *Reading* < *Local* < *French* < *Couple* (from slowest to fastest). In this case there doesn't seem to be a strong effect of the origin of the strangers, the rhythmic difference being mostly visible between speech addressed to a spouse and speech addressed to strangers. Reading, unsurprisingly, shows the slowest speech rate. These speech rate differences are illustrated in Figure 11 and Table 9.

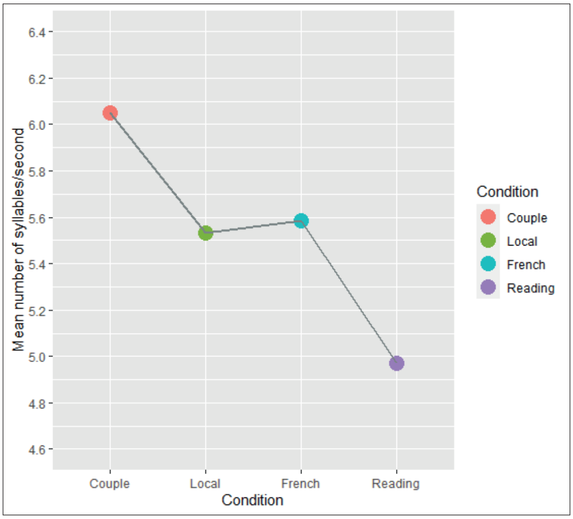

Figure 11: Mean speech rate (nb of syllables/ second in each IPU) for each condition of production (all speakers and vowel classes pooled)

Condition	Mean Rate	Sd Rate
Couple	6.09	2.61
Local	5.53	2.51
French	5.57	2.59
Reading	4.84	1.59

Table 9: Mean and standard deviation values of speech rate for each condition of production (in nb of syll/sec for each IPU)

Regarding the interaction between Sex and Condition, the post-hoc tests showed that for women all conditions were significantly different ($p<0.001$), *Couple* showed the fastest rate, *French* was a little slower, followed by *Local*, and *Reading* was associated with the slowest rate. In men's speech *French* and *Local* showed the same speech rate, while the rest of the pattern was identical: *Couple* > *Local* ≥ *French* > *Reading* ($p<0.001$). Overall women tend to speak a little faster (but not significantly), except for the *Local* condition (see Figure 12 and Table 10). This difference in speech addressed to the two experimenters is interesting, but difficult to interpret in the context of this study; some thoughts will however be offered in the conclusion.

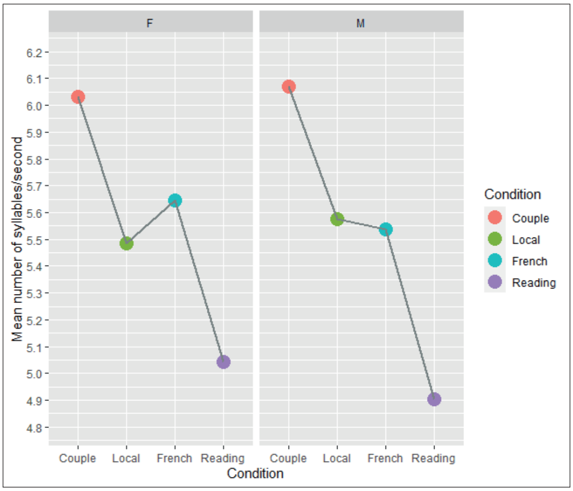

Figure 12: Mean speech rate (nb of syll/second for each IPU) in each production condition for men (right) and women (left)

Sex	Condition	Mean Rate	Sd Rate
F	Couple	6.09	2.63
M	Couple	6.09	2.59
F	Local	5.51	2.57
M	Local	5.56	2.46
F	French	5.64	2.59
M	French	5.51	2.59
F	Reading	4.91	1.63
M	Reading	4.78	1.56

Table 10: Mean and standard deviation values of speech rate for each condition, for men and women (in nb of syll/second for each IPU)

5. Conclusions

In this paper we addressed the issue of social distance and its instantiation through hypo-hyperarticulation. We reported analyses and results on a subpart of a corpus recorded using the Diapix task. Twenty speakers (10M/10F) of Quebec French performed the task with their spouse, with a local stranger, and with a French speaker from France. The three interlocutors aimed at roughly reconstructing a scale of social distance: the spouse represents the closest relationship, followed by the local stranger (with a common cultural background); the speaker from France instantiates the most distant relationship. A reading control task completes the protocol. A total of 109 134 vowel tokens pertaining to 16 vowel classes specific to Quebec French were analyzed, the degree of spectral and temporal reduction of speech being measured through 5 metrics: vowel duration (s), vowel distance to the centroid of the speaker's system, vowel distance to the centroid of the class (also speaker-dependent), compacity (F2-F1 in z-scores), and speech rate (nb of syll/sec in each IPU). If the research question of the effect of social distance on speech is not in itself new, it is now addressed with: 1) a more clearly defined notion of social distance; 2) an appropriate protocol that favors ecological speech while maintaining constant all other elements of the communication situation; 3) a quantity of data that far exceeds what is found in previous studies on the topic.

Table 11 summarizes the results, focusing on the effect of social distance (called Condition in our set up) and the interaction between social distance and the speaker's sex.

Effect	**Metrics**	**Patterns**
Condition	Duration	*Couple* < *Local* = *French* < *Reading* (p<0.001)
	DistSyst	*Couple* < *Local* < *French* < *Reading* (p<0.005)
	DistClass	*Couple* = *Local* ; *Local* = *French* ; *Couple* < *French* ; *Couple, Local, French* < *Reading* (p ≤ 0.05)
	Compacity	*Couple* = *Local* < *French* < *Reading* (p<0.001)
	Speech rate	*Reading* < *Local* < *French* < *Couple* (p<0.005)
Condition: Sex	Duration	_
	DistSyst	♂ *Local* < *Couple* < *French* < *Reading* (p<0.01) ♀ *Couple* < *Local* < *French* < *Reading* (p<0.001)
	DistClass	♂ *Couple* = *French* ; *Local* = *French*; *Local* < *Couple*; *Couple* = *Local* =*French* < *Reading* (p<0.05) ♀ *Couple* = *Local* =*French* < *Reading* (p<0.001)
	Compacity	♂ *Couple* = *Local* < *French* < *Reading* (p<0.001) ♀ *Couple* = *Local* < *French* < *Reading* (p<0.001)
	Speech rate	♂ *Couple* > *Local* = *French* > *Reading* (p<0.001) ♀ *Couple* > *French* > *Local* > *Reading* (p<0.001)

Table 11: Summary of the patterns observed between social distance and the metrics analyzed for the effect of Condition and its interaction with Sex[16]

In a nutshell, speakers hypoarticulate more when interacting than when reading, and when interacting with their spouse than with strangers (and sometimes more with a local stranger than with a foreigner). Speakers therefore partly convey social distance by their position on the hyper-hypo continuum. It

[16] The p-values come from the post-hoc tests presented above. DistSyst stands for "distance to the centroid of the system" and DistClass stands for "distance to the centroid of the class". "_" indicates that the metric was not considered.

now remains to be determined whether there is a convergence/divergence of these measures between speakers over the course of interactions, and to what extent these differences are perceived by listeners.

It appears that the results on the distance to the centroid of the class were not consistent with the other results. Vowel classes were more homogeneous in *Couple* than in *Reading*; this could be interpreted as more hypoarticulated in *Reading* than in *Couple*, which seems counterintuitive. The fact that this result goes against those on our 4 other metrics leads us to think that DistClass might not be a good indicator of the link between social distance and hypoarticulation. Thus, we feel that this measure should be left aside to answer our research questions.

We also observe differences between men and women. For women and contrary to men, the size of the vowel space varied as predicted by our hypothesis. But women unexpectedly tend to speak faster with the *French* experimenter than with the *Local* one. One possible explanation could involve the experimenters' sex: since *French* was a woman and *Local* a man, women might have spoken faster with a fellow woman than with a man. Another interpretation could exploit the idea that women use more standard forms than men (Labov 1990), and might thus be expected to accommodate more than men to the French experimenter, who is most likely perceived both as standard and a fast speaker. This difference in the behavior of men and women would be consistent with Namy *et al.* (2002), who showed in a study on convergence that women accommodate more than men. To further investigate this phenomenon, we reran the statistic models and switched Condition for Interlocutor Sex. We won't detail those results here, but the comparison between the models seems to point toward an accommodation according to regional origin for women in *French* and an accommodation according to social distance for men in *French*.

The definition of our vowel classes also warrants further discussion. On the one hand, we were led to group in a single class vowels appearing in different segmental and syllabic contexts, for instance [i] in open syllables and in syllables closed by [ʁ] and [v, z, ʒ]. The validity of such groupings remains to be confirmed. On the other hand, and more importantly, we could not take into account the prosodic status of the vowels, depending on position in the word (i.e. in word-final *vs* non-final syllables) and degree of stress. But these factors are known to affect the quality and duration of vowels and our vowel classes are not homogeneously distributed with respect to them (for instance, A# appears mostly in word-final position, whereas O# is banned from this position). Thus, prosodic position should be considered in future investigations.

References

Adda-Decker, M., Lamel, L. (1999), "Pronunciation variants across system configuration, language and speaking style", *Speech Communication*, 29/2-4, p. 83-98.

Adank, P., Smits, R., Van Hout, R. (2004), "A comparison of vowel normalization procedures for language variation research", *Journal of the Acoustical Society of America*, 116/5, p. 3099-3107.

Aubanel, V., Nguyen, N. (2020), "Speaking to a common tune: between-speaker convergence in voice fundamental frequency in a joint speech production task", *PloS one*, 15/5, e0232209.

Audibert, N., Fougeron, C., Gendrot, C., Adda-Decker, M. (2015), "Duration- vs. style-dependent vowel variation: a multiparametric investigation", in *18th International Congress of Phonetic Sciences* (ICPhS'15), p. 5.

Baker, R., Hazan, V. (2011), "Diapix Uk: task materials for the elicitation of multiple spontaneous speech dialogs", *Behavior research methods*, 43/3, p. 761-770.

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015), "Fitting linear mixed-effects models using lme4", *Journal of Statistical Software*, 67/1, p. 1-48.

Becker, H., Useem, R.-H. (1942), "Sociological analysis of the dyad", *American Sociological Review*, 7/1, p. 13-26.

Bigi, B. (2012), "SPPAS: a tool for the phonetic segmentations of speech", in *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, European Language Resources Association, Istanbul, p. 1748-1755.

Brown, P., Fraser, C. (1979), "Speech as a marker of situation", in Scherer, K. R., Giles, H. (eds), *Social markers in speech*, Cambridge University Press, Cambridge, p. 33-62.

Chalier, M. (2019), *Les normes de prononciation du français: une étude perceptive panfrancophone*, Ph D thesis, University of Vienna.

Chalier, M. (2021), *Les normes de prononciation du français: une étude perceptive panfrancophone*, De Gruyter, Berlin.

Côté, M.-H. (2012), "Laurentian French (Quebec): extra vowels, missing schwas and surprising liaison consonants", in Gess R., Lyche C., Meisenburg T. (eds), *Phonological variation in French: Illustrations from three continents*, John Benjamins, Amsterdam/Philadelphia, p. 235-274.

Delvaux, V., Soquet, A. (2007), "The influence of ambient speech on adult speech productions through unintentional imitation", *Phonetica*, 64/2-3, p. 145-173.

Ernestus, M., Hanique, I., Verboom, E. (2015), "The effect of speech situation on the occurrence of reduced word pronunciation variants", *Journal of Phonetics*, 48, p. 60-75.

Fabricius, A., Watt, D., Johnson, D. E. (2009), "A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics", *Language Variation and Change*, 21/3, p. 413-435.

Flynn, N., Foulkes, P. (2011), "Comparing vowel formant normalization methods", in *Proceedings of the 16th International Congress of Phonetic Sciences*, p. 683-686.

Gasiorek, J., Giles, H., Soliz, J. (2015), "Accommodating new vistas", *Language & Communication*, 41, p. 1-5.

Gendrot, C., Adda-Decker, M. (2005), "Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German", in *Proceedings of Interspeech 2005*, p. 2453-2456.
Gendrot, C., Adda-Decker, M. (2007), "Impact of duration and vowel inventory size on formant values of oral vowels: an automated formant analysis from eight languages", in *Proceedings of the 16th international congress of phonetic sciences*, Institute of Phonetics, Saarland University Saarbrüken, Germany, p. 1417-1420.
Giles, H., Smith, P. (1979), "Accommodation Theory: optimal levels of convergence", in Giles, H., Clair, St. (eds), *Language and social psychology*, Basil Blackwell, Oxford, p. 45-65.
Giles, H., Ogay, T. (2007), "Communication accommodation theory", in Whaley, B. B., Samter, W. (eds), *Explaining communication: contemporary theories and exemplars*, Lawrence Erlbaum, Mahwah, p. 293-310.
Goldman, J.-P., François, T., Roekhaut, S., Simon, A.-C. (2010), « Étude statistique de la durée pausale dans différents styles de parole », in *Proceedings of the "Journées d'Etude sur la Parole"*, p. 161-164.
Gregory, S. W., Hoyt, B. R. (1982), "Conversation partner mutual adaptation as demonstrated by Fourier series analysis", *Journal of Psycholinguistic Research*, 11/1, p. 35-46.
Gregory, S. W. (1990), "Analysis of fundamental frequency reveals covariation in interview partners' speech", *Journal of Nonverbal Behavior*, 14/4, p. 237-251.
Gregory, S. W., Webster, S. (1996), "A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions", *Journal of Personality and Social Psychology*, 70/6, p. 1231.
Gregory, S. W., Dagan, K., Webster, S. (1997), "Evaluating the relation of vocal accommodation in conversation partners' fundamental frequencies to perceptions of communication quality", *Journal of Nonverbal Behavior*, 21/1, p. 23-43.
Gregory, S. W., Green, B. E., Carrothers, R. M., Dagan, K. A., Webster, S. (2001), "Verifying the primacy of voice fundamental frequency in social status accommodation", *Language & Communication,* 21/1, p. 37-60.
Grosman, I., Simon, A.-C., Degand, L. (2018), « Variation de la durée des pauses silencieuses: impact de la syntaxe, du style de parole et des disfluences », *Langages*, 211/3, p. 13-40.
Harmegnies, B., Poch-Olivé, D. (1992), "A study of style-induced vowel variability: laboratory versus spontaneous speech in Spanish", *Speech Communication*, 11/4-5, p. 429-437.
Harmegnies, B., Poch-Olivé, D. (1994), "Formants frequencies variability in French vowels under the effect of various speaking styles", *Le Journal de Physique IV,* 4/C5, p. 509-512.
Huet, K., Harmegnies, B. (2000), « Contribution à la quantification du degré d'organisation des systèmes vocaliques », in *Proceedings of the "Journées d'Étude sur la Parole"*, p. 225-228.
Hupin, B., Simon, A.-C. (2007), « Analyse phonostylistique du discours radiophonique. Expériences sur la mise en fonction professionnelle du phonostyle et sur le lien entre mélodicité et proximité du discours radiophonique », *Recherches en communication*, 28, p. 103-122.

Kelley, H., Berscheid, E., Christensen, A., Harvey, J., Huston, T., Levinger, G., McClintock, E., Peplau, L., Peterson, D. (1983), *Close relationships*, W. H. Freeman, New York.

Koppen, K., Ernestus, M., van Mulken, M. (2017), "The influence of social distance on speech behavior: Formality variation in casual speech", *Corpus Linguistics and Linguistic Theory*, 15/1, p. 139-165.

Labov, W. (1990), "The intersection of sex and social class in the course of linguistic change", *Language Variation and Change*, 2, p. 205-54.

Lancien, M., Côté, M.-H., Bigi, B. (2020), "Developing resources for automated speech processing of Quebec French", in *Proceedings of the 12th Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, p. 5325-5330.

Lancien, M. (2021), *Le rôle de la réduction phonétique dans l'expression de la distance sociale: étude acoustique des voyelles orales du français québécois dans différentes situations de communication*, Ph D thesis, Université de Lausanne.

Lê, S., Josse, J., Husson, F. (2008), "FactoMineR: a package for multivariate analysis", *Journal of Statistical Software*, 25/1, p. 1-18, doi: 10.18637/jss.v025.i01.

Lenth, R. (2020), "emmeans: Estimated Marginal Means, aka Least-Squares Means", R package version 1.5.3, https://CRAN.R-project.org/package=emmeans.

Léon, M., Léon, P. (1997), *La prononciation du français*, Nathan, Paris.

Lindblom, B. (1990), "Explaining phonetic variation: a sketch of the H&H theory", in Hardcastle, W. J., Marchal, A. (eds), *Speech production and speech modelling*, Kluwer, Dordrecht, p. 403-439.

Lobanov, B. M. (1971), "Classification of Russian vowels spoken by different speakers", *Journal of the Acoustical Society of America*, 49/2B, p. 606-608.

Lyche, C. (2010), « Le français de référence: éléments de synthèse », in Detey, S. *et al.* (eds), *Les variétés du français parlé dans l'espace francophone: ressources pour l'enseignement*, Ophrys, Paris, p. 143-165.

Maisonneuve, J. (1966), *Psychosociologie des affinités*, Ph D thesis, Université de Paris.

Martin, P. (1998), « Dynamique vocalique en français du Québec », *La Linguistique*, 34/2, p. 67-76.

Martin, P. (2002), « Le système vocalique du français du Québec: de l'acoustique à la phonologie », *La linguistique*, 38/2, p. 71-88.

Namy, L. L., Nygaard, L. C., Sauerteig, D. (2002), "Gender differences in vocal accommodation: the role of perception", *Journal of Language and Social Psychology*, 21/4, p. 422-432.

Paradis, C. (1985), *An acoustic study of the variation and change in the vowel system of Chicoutimi and Jonquiere (Quebec, CANADA)*, Ph D thesis, Université du Québec à Chicoutimi.

Pardo, J. (2006), "On phonetic convergence during conversational interaction", *Journal of the Acoustical Society of America*, 119/4, p. 2382-2393.

Pardo, J., Gibbons, R., Suppes, A., Krauss, R. M. (2012), "Phonetic convergence in college roommates", *Journal of Phonetics*, 40/1, p. 190-197.

Pardo, J., Urmanche, A., Wilman, S., Wiener, J. (2017), "Phonetic convergence across multiple measures and model talkers", *Attention, Perception & Psychophysics*, 79/2, p. 637-659.

Pardo, J., Urmanche, A., Wilman, S., Wiener, J., Mason, N., Francis, K., Ward, M. (2018), "A comparison of phonetic convergence in conversational interaction and speech shadowing", *Journal of Phonetics*, 69, p. 1-11.

Pardo, J., Urmanche, A., Gash, H., Wiener, J., Mason, N., Wilman, S., Francis, K., Decker, A. (2019), "The montclair map task: Balance, efficacy, and efficiency in conversational interaction", *Language and Speech*, 62/2, p. 378-398.

Poch-Olivé, D., Harmegnies, B. (1992), « Variations structurelles des systèmes vocaliques en français et espagnol sous l'effet du style de parole », *Journal de Physique IV*, 2/C1, p. 283-286.

R CORE TEAM (2018), "R: A Language and Environment for Statistical Computing", R Foundation for Statistical Computing, Vienna, Austria.

Remysen, W. (2020), « S'adapter pour se sentir moins insécurisé ? L'(in)sécurité linguistique au Québec sous l'angle de l'accommodation », in Lorilleux J., Feussi V. (éds), *L'(in)sécurité linguistique en Francophonies: perspectives interdisciplinaires*, L'Harmattan, Paris, p. 63-75.

Rouas, J.-L., Beppu, M., Adda-Decker, M. (2010), "Comparison of spectral properties of read, prepared and casual speech in French", in *Proceedings of the 7th Language Resources and Evaluation Conference*, European Language Resources Association, Valletta, p.606-611.

Santerre, L. (1974), « Deux *e* et deux *a* phonologiques en français québécois: étude phonologique, articulatoire et acoustique des oppositions de timbre et de durée », *Cahier de linguistique*, 4, p. 117-145.

Santerre, L. (1976), « La postériorisation des *a* en français montréalais », in *Rapport au Conseil des Arts*.

Santerre, L. (1979), « La comparaison des /ɛ/ et des /a/ en québécois et en français », in Rondeau, G. *et al.* (éds), *Vingt-cinq ans de linguistique au Canada: hommage à Jean-Paul Vinay*, CEC, Montréal, p. 325-361.

Scarborough, R., Dmitrieva, O., Hall-Lew, L., Zhao, Y., Brenier, J. (2007), "An acoustic study of real and imagined foreigner-directed speech", *Journal of the Acoustical Society of America*, 121/5, p. 3044.

Scarborough, R., Zellou, G. (2013), "Clarity in communication: 'clear' speech authenticity and lexical neighborhood density effects in speech production and perception", *Journal of the Acoustical Society of America*, 134/5, p. 3793-3807.

Séguin, M.-C. (2010), *Catégorisation acoustique des voyelles moyennes antérieures arrondies en français laurentien*, Ph D thesis, University of Ottawa.

Sievert, C. (2020), "Interactive web-based data visualization with R, plotly, and shiny", Chapman and Hall/CRC Florida, 2020.

Simon, A.-C., Auchlin, A., Avanzi, M., Goldman, J. P. (2010), « Les phonostyles: une description prosodique des styles de parole en français », in Abecassi, M., Ledegen, G. (éds), *Les voix des Français. En parlant, en écrivant*, Peter Lang, Bern, p. 71-88.

Wickham, H. (2016), *ggplot2: Elegant graphics for data analysis*, Springer, New York.

Walker, D. C. (1984), *The pronunciation of Canadian French*, University of Ottawa Press, Ottawa.

Yuan, J., Liberman, M., Cieri, C. (2006), "Towards an integrated understanding of speaking rate in conversation", in *Proceedings of the 9th International Conference on Spoken Language Processing*.