ACOUSTIC DESCRIPTION OF QUEBEC FRENCH HIGH VOWELS: FIRST RESULTS

Vincent Arnaud[a], *Caroline Sigouin*[a] *& Johanna-Pascale Roy*[b]

[a]Université du Québec à Chicoutimi, Canda; [b]Université Laval, Canada
Vincent.Arnaud@uqac.ca

ABSTRACT

Vowels are traditionally described as static acoustic targets. This study evaluates the impact of spectral changes on the description of Quebec French tense/lax high vowels. Tense variants are expected in open syllables while lax variants are observed in syllables closed by a non-lengthening consonant. We argue that these contextual variants are more accurately described as trajectories through acoustic space.

Keywords: high vowels, Quebec French, VISC, contextual variants

1. INTRODUCTION

This contribution focuses on the acoustical analysis of vocalic tenseness variation in the vowel system of Quebec French (QF). The QF vowel system includes three high vowels /iyu/ that are characterized by a variability of their phonetic realizations. The main source of variability is the contextual variation between tense vowels [iyu] and lax vowels [ɪʏʊ]. Tense variants are produced in open syllables and are also expected in syllables closed by lengthening consonants /ʁvzʒ/. Conversely, lax variants [ɪʏʊ] occur in syllables closed by non-lengthening consonants. This rule is compulsory in stressed closed syllables. The laxing of high vowels in this context is recognized as one of the major characteristics of QF. This phenomenon presents no specific correlation with social factors or with the degree of formality. There seems to be no negative perception of the lax variants by the QF community, if they are perceived at all [15].

To our knowledge, the literature about the acoustic differences between these contextual variants is relatively scarce and limited to the analysis of duration and central frequency of the first two formants (F_1, F_2). Paradis [14] has acoustically described the vowel system of QF spoken by speakers from Saguenay in the eastern part of the province of Quebec. The data was extracted from sociolinguistic interviews. Dolbec, *et al.* [4] have analyzed high vowels produced by QF speakers reading monosyllabic words included in carrier sentences. Martin [10] focused on vocalic occurrences extracted from isolated words read by speakers from different regions of Quebec. These three studies noted a peripherality of the tense variants in open syllables and a centralization of the lax variants [ɪʏʊ] in a $F_1 \times F_2$ space. This centralization was reflected by higher F_1 values, lower F_2 values for front vowels [ɪʏ] and higher F_2 values for the back vowel [ʊ]. Martin [10] also remarked that the duration of the lax realizations represents only 80-85% of the duration of their tense counterparts. MacLeod, *et al.* [9] compared the production of high vowels [iɪuʊ] in Canadian English and Canadian French produced by early bilingual and monolingual speakers. The French monolinguals produced the vowels labeled [ɪ] and [ʊ] with lower F_1 values than the English monolinguals. With the exception of [ʊ], the French monolinguals produced vowels with more extreme F_2 values than the English monolinguals. Therefore, the French high vowels showed a higher degree of peripherality in a $F_1 \times F_2$ space.

2. THEORETICAL ASPECTS

2.1. Beyond F_1-F_2

F_1 and F_2 appear to be the main acoustic correlates of vowel identity. Thus, in the various studies cited above, neither F_3 nor *f*0 were taken into account. However, F_2 and F_3 are efficient candidates for indicating the contrast between [i] and [y] in French [16]. Some speakers showed a large decrease in F_3 and little or no F_2 variation; others showed a large decrease in F_2 and F_3. At a perceptual level, these idiosyncratic acoustic strategies could be interpreted as expressions of F'_2, a non-linear integration of F_2 and higher formants [3]. Moreover, in the framework of the Dispersion-Focalization Theory [17], [i], [y] and [u] are considered to be representative prototypes of the focalization process with the vicinity of two of their formants. More recently, Gendrot, *et al.* [5]

have suggested a higher degree of focalization of [i] and [y] in French than in seven other languages. In acoustic terms, this high degree of focalization was related to a small F_4-F_3 or F_3-F_2 distance.

In addition, Traunmüller [18] has shown the strong correlation between perceived vowel height and the value (in bark) of the distance between F_1 and *f*0. This observation has been confirmed for continental French [11].

2.2. Beyond static parameterization

In the studies cited above, formant measurements were made at "steady-state" time. Nevertheless, reducing the acoustic portrayal of the vowels to a static parameterization has important limitations. The vowel inherent spectral change (VISC, [13]) has been found to be an important factor in North American English vowel identification [8, 13]. For instance, Nearey and Assmann [13] have collected F_1 and F_2 values measured at the beginning and the end of 10 different Western Canadian vowels spoken in isolation. They concluded that a large majority of vowels show spectral movement in the $F_1 \times F_2$ space. Hillenbrand, *et al.* [8] analyzed *f*0, F_1, F_2, F_3 and duration of American English vowels in /hVd/ syllables. By using quadratic discriminant analysis (QDA), this study points out that vowel classification accuracy is improved from a parameterization including a unique sample of the formant pattern taken at the "steady state" to two samples taken at 20% and 80% of the vowel duration. Adding a third data sample extracted at 50% of the vowel duration produced no additional benefit. Other studies [7, 12] confirm that spectral information conveyed in vowel onset and offset has a major impact on the North American English vowel identification and that these vowels are more accurately described as trajectories through acoustic space and not as static targets.

If the VISC appears to be relevant for North American English vowel identification, three models dedicated to its acoustic parameterization are competing [12, 13]. All underline the importance of a sample of the formant pattern near the onset. However, they disagree about the parametric representation of formant trajectories. The *dual target* hypothesis (DTH) considers that two explicit samples of the formant pattern, one near the onset and one near the offset of the vowel, are relevant. The *target plus slope* hypothesis (TSH) claims that the relevant cue is the rate of change of formant frequencies over time. The *target plus direction* hypothesis (TDH) argues that only the direction of change in formant frequencies is important. In their recent study, Morrisson and Nearey [12] support the efficiency of the DTH in the identification of synthetic English vocalic stimuli by Western Canadian English listeners.

2.3. Objectives

In this contribution dedicated to the acoustic description of QF tense/lax high vowels, three hypotheses will be developed:

1. To estimate the role of *f*0 and F_3 in the classification accuracy of the QF tense high vowels and their lax counterparts;
2. To investigate how the VISC could contribute to better understanding of the variation between these six QF variants;
3. To evaluate the potential differences induced by the three hypotheses (cf. Section 2.2) about the parameterization of the VISC.

3. METHODOLOGY

The subjects were twelve undergraduate native QF speakers (six men, six women), aged from 20 to 29, from the town of Saguenay. A set of six nonsense French words was constructed by commuting /i/, /y/ and /u/ in two different consonantal contexts: /bV/ and /bVb/. Because of the laxing rule depicted earlier, the tense variants [iyu] were expected in open syllables while lax variants [ɪʏʊ] were expected in closed syllables.

These monosyllabic nonsense words were embedded in a carrier sentence which had the following generic structure: *C'est des* ***CV*** *(or)* ***CVC*** *que je dis.* (cf. "audio file 1"). Some carrier sentences containing distracters were added. Ten repetitions of each target sequence were produced by the subjects. The corpus was presented in random order. The speakers were asked to produce a prosodic contrastive focus on each target sequence. Each speaker has been recorded in an anechoic room with an Audio Technica AT831 microphone connected to a Tascam HD P2 digital recorder (44.1 kHz, 16 bits, mono).

A total of 718 high vowel tokens were retained. The final number of repetitions in each category and for each speaker varied between 5 and 13, rather than the expected 10. Some tokens could not be analyzed (creaky voice, devoicing…). Likewise, some supplemental iterations were added to our sample. The tokens were labeled manually using *Praat* [2]. For each token, we were interested in measuring vowel duration and values of *f*0, F_1, F_2 and F_3 at 25%, 50%, and 75% duration.

For the acoustic parameterization of the VISC, different variables were considered:

- DTH: **Two targets** were represented by four different sets of parameters combining *f*0, F_1, F_2, F_3 sampled at 25% and 75% of the vowel duration.
- TSH: Four variables labeled S_0, S_1, S_2 and S_3 were used to express the **rate of change** of *f*0, F_1, F_2 and F_3 values (ΔF) from 25% (t_{25}) to 75% (t_{75}) of the vowel duration. This is expressed as $\Delta F/(t_{75}-t_{25})$.
- TDH: Two to four variables were used to express the **direction of change** as the ratio between the change of *f*0, F_1, F_2 and F_3 values (ΔF) from 25% to 75% of the vowel duration and the "length" of a vector defined in four n-dimensional Euclidean spaces (2D space *a*: $F_1 \times F_2$; 3D space *b*: $f0 \times F_1 \times F_2$; 3D space *c*: $F_1 \times F_2 \times F_3$; 4D space *d*: $f0 \times F_1 \times F_2 \times F_3$). These variables are labeled C_{0y}, C_{1y}, C_{2y} and C_{3y} where the number indicates which parameter is concerned (*0*=*f*0, *1*=F_1…) and *y* indicates one of the four preceding n-dimensional spaces.

4. RESULTS

Tokens were grouped depending on the expected phonetic quality (tense/lax) into six target categories labelled *bi#*, *bib*, *by#*, *byb*, *bu#*, *bub*.

Image file 1 shows the repartition of F_1 and F_2 values at 50% of the vowel duration. A relatively clear graphical distinction emerges between the six categories. However, the pairs [uʊ] and [ɪy] show an overlap in this static representation. Nevertheless, as shown in image files 2 and 3 (which represent the dispersion of the tokens in two distinct $F_1 \times F_2$ spaces: the first at 25% and the second at 75% of the vowel duration), lax and tense variants exhibit opposite spectral trajectories. Between the first and the second quarter of their production, tense variants seem to increase their degree of peripherality while lax variants appear to show an increase in centralization.

To examine more precisely the influence of *f*0, F_3 and spectral change information on the degree of distinction of vowel categories, a series of QDAs ("jackknife" method) was carried out. In QDAs, log-transformed frequency values are used rather than the raw Hertz values [12, 13]. Table 1 illustrates the effects of DTH on the percentage of correctly classified vowel tokens. The one-sample results are based on four combinations of acoustic parameters at 50% of the vowel duration; the two-sample results are based on samples taken at 25% and 75% of the vowel duration; the three-sample results are based on samples taken at 25%, 50%, and 75% of the vowel duration. (The label "Dur." indicates that the duration was included).

Table 1: QDA results (in percent) showing the effect of including duration and spectral change information on classification accuracy. Percentages in parentheses are QDA results obtained using raw Hertz values.

	50%		25-75%		25-50-75%	
	No dur.	Dur.	No dur.	Dur.	No dur.	Dur.
F_1, F_2	84 (83)	87 (87)	90 (90)	92 (91)	90 (90)	92 (91)
F_1, F_2, F_3	95 (95)	96 (95)	98 (98)	**99 (99)**	98 (98)	**99 (99)**
*f*0, F_1, F_2	93 (92)	94 (93)	96 (96)	96 (95)	95 (94)	95 (94)
*f*0, F_1, F_2, F_3	97 (97)	98 (97)	98 (98)	**99 (99)**	**99 (99)**	**99 (99)**

Including F_3 in the acoustic parameter set has a major effect on the classification accuracy. However, the inclusion of spectral change information seems to have a more consistent effect on the classification accuracy than including the duration and/or *f*0. No matter the combination of acoustic parameters, the classification accuracy increases when comparing a single sample with two samples. Nevertheless, as shown in English [8], adding a third sample does not appear to improve classification accuracy. Furthermore, as shown in Dutch [1], the use of a log-transformed scale does not seem to affect classification accuracy.

Table 2 illustrates the effects of TSH and TDH on the degree of distinction of vowel categories.

Table 2: QDA results showing the effect of TSH and TDH on classification accuracy. These results are based on log-transformed values of *f*0, F_1, F_2, F_3.

	Acoustic parameters set	Percentage
TSH	F_1, F_2, S_1, S_2	91
	*f*0, F_1, F_2, S_0, S_1, S_2	94
	F_1, F_2, F_3, S_1, S_2, S_3	**98**
	*f*0, F_1, F_2, F_3, S_0, S_1, S_2, S_3	98
TDH	F_1, F_2, C_{1a}, C_{2a}	89
	*f*0, F_1, F_2, C_{0b}, C_{1b}, C_{2b}	94
	F_1, F_2, F_3, C_{1c}, C_{2c}, C_{3c}	96
	*f*0, F_1, F_2, F_3, C_{0d}, C_{1d}, C_{2d}, C_{3d}	**98**

These results take into account different combinations of acoustic parameters at 25% of the vowel duration and various variables (cf. Section 3) representing the rate of change or the direction of change of each acoustic parameter. Again, the impact of F_3 on classification accuracy is highlighted, but the effect of *f*0 seems to be negligible. As observed in English [13], the classifications based on TDH seem to be slightly less accurate than the classifications based on DTH and TSH when the number of acoustic parameters included in QDA decreased.

5. DISCUSSION AND CONCLUSION

Two problems arise from our study. First, because QF tense and lax variants appear in different contexts, we cannot distinguish the contributions of the VISC and coarticulation on the spectral trajectories of the tokens studied. Second, the pattern of durational differences among the contexts is totally reversed to that observed by Martin [10]. As shown in Table 3 the duration of the tense realizations represents only 59-63% of the duration of their lax counterparts. The effect of the sex of the speaker on duration is not significant. However, a Kruskal-Wallis test (H(5)=49.181, p<.01) points out significant differences between the mean duration of the tokens produced in open syllables and the mean duration of the tokens produced in closed syllables.

Table 3: Average duration of high QF vowels produced by six men (M) and six women (W).

	Sex	bi#	bib	bu#	bub	by#	byb
Duration (s)	W	0.119	0.203	0.126	0.192	0.137	0.217
	M	0.111	0.183	0.121	0.194	0.120	0.202

This undocumented result could be related to the voiced consonantal context. In English, the vowels produced in CVC are systematically longer when they are preceded and followed by voiced stops rather than unvoiced stops [6]. This duration pattern could also be linked to the geographic origin of the speakers. In-progress analyses suggest a potential distinct duration pattern between Quebec City and Saguenay. Nevertheless, the small effect of the duration on QDA classification accuracy leads us to ask ourselves if, as shown in English [6], *human* QF listeners would give little weight to modifications of duration in the distinction of the QF tense/lax vowels. Despite those remaining questions, this contribution seems to confirm the tendencies observed in English [7, 8, 12, 13] and argues for taking into account F_3 [5, 16, 17] and spectral changes in the description of French vowels.

6. ACKNOWLEDGEMENTS

This project is supported by the Fonds Québécois de Recherche sur la Société et la Culture and the Fonds institutionnel de recherche de l'UQAC. We are grateful to Mathieu Cusson for his help with statistics and to Elizabeth Winokur and Leif French for proofreading the manuscript.

7. REFERENCES

[1] Adank, P., Smits, R., van Hout, R. 2004. A comparison of vowel normalization procedures for language variation research. *J. Acoust. Soc. Am.* 116, 3099-3107.

[2] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5, 341-345.

[3] Carlson, R., Granström, B., Fant, G. 1970. Some studies concerning perception of isolated vowels. *STL-QPSR* 2-3, 19-35.

[4] Dolbec, J., Ouellon, C., Ouellet, M. 1993. Considérations préliminaires sur le trapèze vocalique de locuteurs québécois. *Dialangue* 4, 44-48.

[5] Gendrot, C., Adda-Decker, M., Vaissière, J. 2008. Les voyelles /i/ et /y/ du français: Focalisation et variations formantiques. *Proc. 27èmes JEP* Avignon, 205-208.

[6] Hillenbrand, J.M., Clark, M.J., Houde, R.A. 2000. Some effects of duration on vowel recognition. *J. Acoust. Soc. Am.* 108, 3013-3022.

[7] Hillenbrand, J.M., Clark, M.J., Nearey, T. 2001. Effects of consonant environment on vowel formant patterns. *J. Acoust. Soc. Am.* 109, 748-763.

[8] Hillenbrand, J.M., Getty, L., Clark, M.J., Wheeler, K. 1995. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 97, 3099-3111.

[9] MacLeod, A., Stoel-Gammon, C., Wassink, A. 2009. Production of high vowels in Canadian English and Canadian French: A comparison of early bilingual and monolingual speakers. *J. Phonetics* 37, 374-387.

[10] Martin, P. 2002. Le système vocalique du français du Québec. De l'acoustique à la phonologie. *La Linguistique* 38, 71-88.

[11] Ménard, L., Schwartz, J.L., Boë, L.J., Kandel, S., Vallée, N. 2002. Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood. *J. Acoust. Soc. Am.* 111, 1892-1905.

[12] Morrison, G., Nearey, T. 2007. Testing theories of vowel inherent spectral change. *J. Acoust. Soc. Am.* 122, EL15-EL22.

[13] Nearey, T., Assmann, P. 1986. Modeling the role of inherent spectral change in vowel identification. *J. Acoust. Soc. Am.* 80, 1297-1308.

[14] Paradis, C. 1985. *An Acoustic Study of Variation and Change in the Vowel System of Chicoutimi and Jonquière (Quebec)*. Ph.D. Thesis, University of Pennsylvania.

[15] Paradis, C., Brousseau, M., Dolbec, J. 1993. Variétés linguistiques et intelligibilité: Enjeux sociolinguistiques pour la synthèse de parole. *Revue québécoise de linguistique* 22, 13-36.

[16] Schwartz, J.L., Beautemps, D., Abry, C., Escudier, P. 1993. Inter-individual and cross-linguistic strategies for the production of the [i] vs [y] contrast, *J. Phonetics* 21, 411-425.

[17] Schwartz, J.L., Boë, L.J., Vallée, N., Abry, C. 1997. The dispersion-focalization theory of vowel systems. *J. Phonetics* 25, 255-286.

[18] Traunmüller, H. 1981. Perceptual dimension of openness in vowels. *J. Acoust. Soc. Am.* 69, 1465-1475.