

A Cross-Language Study of Vowel Categorization and Vowel Acoustics: Canadian English versus Canadian French

Paola Escudero* and Linda Polka‡

*Utrecht University, Utrecht, The Netherlands and McGill University, Montreal, Canada

‡ McGill University, Montreal, Canada

E-mail: paola.escudero@let.uu.nl, linda.polka@mcgill.ca

ABSTRACT

We show the perception of Canadian French (CF) vowels by Canadian English (CE) listeners and test a cue-weighting hypothesis to explain the attested assimilation patterns. Five CF vowels, /i, y, u, ε, æ/, and three allophonic variants, [ɪ, ʏ, ʊ], were examined. The listeners completed a native-language identification task with goodness of fit judgments. We found that most French vowels were identified as more than one vowel category in English. Acoustic analyses of the vowel tokens revealed that the multiple mappings occur because the English listeners paid attention to both spectral and durational cues when identifying the French vowels. We claim that English listeners use the cue-weighting strategies of their first language when approaching a foreign-vowel identification task. We predict the specific problems that CE speakers will face when learning to categorize CF vowels and the possible solutions that they should entertain.

1. INTRODUCTION

Adults often experience difficulties when learning to perceive the sounds of a new language. Cross-language research has shown us that perceptual difficulties vary considerably so that some non-native contrasts are readily perceived while others are quite problematic [1]. Moreover, the perceptual problems that arise take different forms. On the one hand, the non-native listener may fail to perceive phonetic differences that are readily apparent to the native listener. On the other hand, the non-native listener may perceive variability in different productions of a non-native phoneme that the native listener does not. For example, non-native listeners have been shown to assimilate the same non-native phoneme to different native categories when it is produced in different speaking styles or phonetic contexts [2].

The present study is concerned with the latter type of problem with respect to vowel perception. It has been claimed that a description of the relevant perceptual cues or dimensions and their relative weightings in the native language are needed to fully explain the language-specific nature of vowel perception [3]. We are interested in exploring the merits of this approach. In the present study, we assess whether listeners' perceptual weighting of temporal and spectral information in their native language impacts their perception of non-native vowels, and we then

explore the implications of this for L2 learning. To do so, we examine the perception of Canadian French (hereafter, CF) vowels by Canadian English (hereafter, CE) listeners along with a comparative acoustic analysis of vowels in these two languages.

This comparison is particularly informative given the evidence suggesting that the types of cues that carry the most perceptual weight differ in English and French. In English, both spectral and durational properties provide important cues to vowel identity [4]. Although vowel duration is often considered a secondary cue in English, English listeners rely on durational cues to differentiate certain vowels, such as tense-lax pairs. As well, duration carries more perceptual weight in English when spectral information is ambiguous. In contrast to English, it has been suggested that French vowel contrasts are cued mainly by spectral properties and that duration carries little, if any, perceptual weight [5]. These findings are supported by research on American English and European French. The present research replicates and extends these findings to Canadian dialects of these languages. We have three specific aims: 1) to describe the assimilation of CF vowels by CE adults, 2) to consider the role of native language cue weighting to explain these assimilation patterns, and 3) to describe the particular problems that CE learners of CF will encounter and the possible solutions that they should entertain.

2. CE VERSUS CF VOWELS

2.1 Vowel selection and description

For this study, we selected a subset of vowels that occur in CF [6], most of which have a phonemic equivalent in CE. Two sets of vowels were selected. The first set includes the high vowels /i/, /u/ and /y/. In CF /i/, /u/ and /y/ each have a lax allophonic variant, [ɪ], [ʊ], and [ʏ] respectively; these allophones occur obligatorily in final closed syllables. In CE, /i/, /u/, /ɪ/ and /ʊ/ occur but not /y/ or /ɻ/. The second set includes the front vowels /i/, /ɪ/, /ε/ and /æ/ which are phonemic in CE and CF, with the exception of CF [ɪ] which is an allophonic variant of /i/.

2.2 Vowel corpus

Both set 1 and set 2 items were recorded by 3 male and 3 female monolingual speakers of CE and CF. For set 1, /i/, [ɪ], /u/, [ʊ], /y/, and [ʏ] productions were recorded in a /bVs/ or /bVz/ context as required to generate both

allophones. For set 1, six tokens (one per talker) of each vowel were selected. For set 2, /i/, [ɪ], /ɛ/ and /æ/ productions were recorded in several different CVC contexts. Eighteen tokens (6 per talker in 3 contexts) were selected for /i/ and [ɪ]; thirty tokens (6 per talker in 6 contexts) were selected for /ɛ/ and /æ/. All vowels were produced in a sentence context. The carrier sentence in English was “Now I say ____ again”; the carrier sentence in French was “Maintenant je dis ____ encore”.

2.3 Acoustic analysis

Spectral and durational properties of the vowels were measured. Formant values (F1 and F2) were derived from LPC analysis in which a 25 ms Hamming window was centered at successive 10 ms intervals across the vowel; the values reported here correspond to the earliest point in the vowel at which the formant values reached a quasi-steady state. Duration values were measured on the waveform from onset to offset of voicing. Mean values for F1, F2, and duration averaged over all tokens of each vowel type (averaging across talkers and contexts) are shown in Figure 1 and 2 below for set 1 (high vowels) and in figure 3 and 4 for set 2 (front vowels).

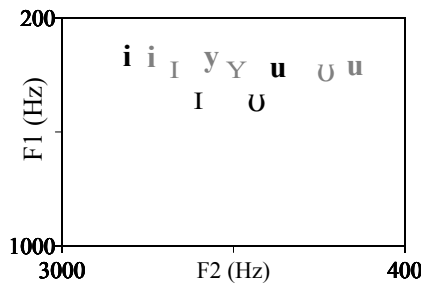


Figure 1: Set 1 High vowels. Mean F1 and F2 frequencies (in Hz); English in **black**; French in **gray**)

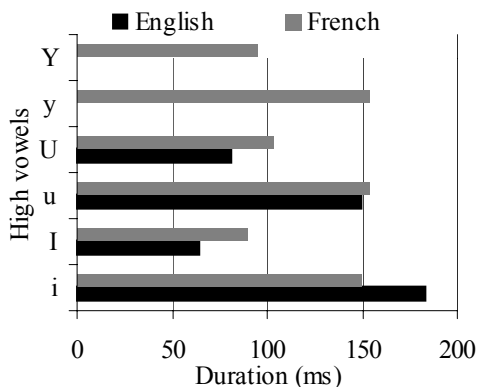


Figure 2: Set 1: High vowels. Mean duration values.

As shown in Figure 1 and 2 above, the CE lax vowels [ɪ], [ʊ] and [ʏ] are shorter and more central than their tense counterparts. CF /i/ and /u/ are close to CE /i/ and /u/ in both spectral and duration properties. CF [ɪ], and [ʊ] are spectrally similar to CE /i/ and /u/, while their duration is similar to CE /ɪ/ and /ʊ/. Both CF /y/ and [ʏ] are similar to English tense vowels with respect to F1 and fall between English front vowels and back vowel with respect to F2.

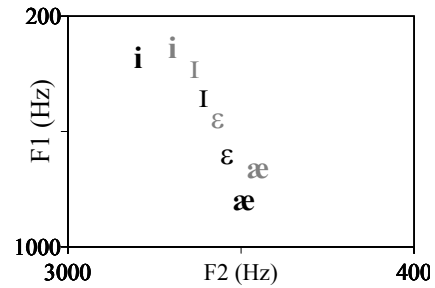


Figure 3: Set 2: Front vowels. Mean F1 and F2 frequencies averaged across consonantal contexts (English in **black**; French in **gray**).

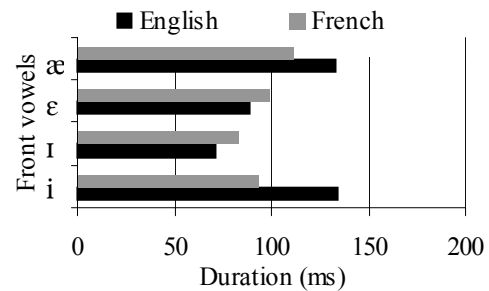


Figure 4: Set 2: Front vowels. Mean duration values.

As shown in Figure 3, all of the CF front vowels are higher (i.e. lower in F1) than their English counterparts. Thus, French /ɪ/ falls between English /i/ and /ɪ/, French /ɛ/ falls between English /ɪ/ and /ɛ/, and French /æ/ falls between English /ɛ/ and /æ/. Figure 4 shows that the duration of French [ɪ] is similar to English /ɪ/, and the duration of French /ɛ/ is similar to English /ɛ/. French /æ/ is shorter than English /æ/ and is closer to English /ɛ/.

3. CE PERCEPTION OF CF VOWELS

Following a cue-weighting view, we expect that the CF vowel segments will be mapped onto more than one English vowel. We expect multiple mappings to occur because CE listeners pay attention to both spectral and durational properties when identifying vowels and because our acoustic analyses reveal that the spectral and duration properties of some CF vowels vary in ways that will pull the English listener toward different native vowel percepts. For instance, because the CF allophonic variants [ɪ], [ʊ] and [ʏ] are spectrally closer to English tense vowels but similar to English lax vowels in duration, English listeners' identification will depend on the relative amount of information provided by each cue. For example, CF /ɪ/ tokens that are long and high will be mapped to English /i/ but CF [ɪ] tokens that are short and low will be labeled as English /ɪ/. CF [ɪ] tokens that are long and low or high and short will not be consistently labeled.

3.1 Cross-language vowel assimilation

Subjects: Fifteen monolingual CE adults identified the CF vowel tokens described above. These listeners had lived in English-speaking regions of Canada for most of their lives

and had lived in Montreal for less than a year at the time of testing.

Task: Subjects completed a native-vowel assimilation task with goodness of fit ratings. For this task, the CVC syllables were extracted from the sentences. Listeners were tested on set 1 items first, and on set 2 items on a different day. They listened to two repetitions of each syllable and were asked to match the vowel in the syllable to one of the options displayed on a computer screen. Right after, the listeners rated the quality of their match on a 5-point scale (1=poor; 5=good). The following seven English keywords, containing the target vowels, were used as response options: “see”, “it”, “moo”, “put”, “pet”, “say”, and “at”.

Results: The labeling results are shown in Table 1 for the high vowels and Table 2 for the front vowels. In the tables, the first column shows the CF vowel that was presented and the next 3 columns show the most frequently chosen category, the overall percent identification and the median rating; the last 3 columns show this same information for the second most frequently chosen response.

CF	Most chosen category			2 nd chosen category		
	CE	% Ident.	Rate	CE	% Ident.	Rate
i	i	91	4	ɪ	8	3
u	u	88	4	ʊ	12	3 -4
y	u	49	3 - 4	ʊ	46	3
ɪ	ɪ	88	4	i	8	3
ʊ	ʊ	72	3 - 4	u	28	3 - 4
ɥ	ʊ	80	3	u	11	3

Table 1: CE identification results for the CF high vowels.

CF						
	CE	% Ident.	Rate	CE	% Ident.	Rate
æ	æ	64	4	ɛ	33	3 -4
ɛ	ɛ	61	4	ɪ	36	4 -5
i	i	75	4	ɪ	24	3 - 4
ɪ	ɪ	80	4 - 5	i	18	4

Table 2: CE identification results for the CF front vowels.

It should be noted that to account for at least 90% of the responses, we must consider assimilation to two native vowel categories; this is the case in both vowel sets and for every CF vowel, except /i/. Moreover, for each of the CF vowels, the ratings assigned to each native vowel assimilation are quite comparable. Thus, English listeners perceived the CF vowels to be equally good instances of two vowels in English. Hence, it appears that these CF vowels reveal a multiple-category assimilation pattern.

As shown in Table 1, English listeners choose an English tense vowel most often when assimilating a CF tense vowel, and they choose an English lax vowel most often when assimilating a CF lax vowel. However, for each CF vowel, there are also some assimilation responses that show the reverse pattern. That is, CF tense vowels are also mapped to an English lax vowel, and some CF lax vowels are also

mapped to an English tense vowel. It is important to note that these multiple assimilations occur for vowels spoken in the same phonetic context and general style; they are not due to variation in syllable context or general elicitation conditions (e.g. citation vs. sentence).

Table 2 shows that overall, the multiple assimilation for CF vowels in set 2 is more distributed across two native vowel categories than in set 1, even for CF /i/ and [ɪ] which appear in both sets. Recall that set 2 includes productions of the vowels in several CVC contexts. However, closer examination of set 2 data reveals that multiple mappings are observed within a phonetic context and therefore this type of variation cannot fully account for this finding. The question now is whether multiple-category assimilation is systematically related to cue variability across tokens; this must be the case if English listeners’ perception involves a native-language cue-weighting filter.

3.2 Cue/response relationships for CF /æ/

As a first step to address the role of cue weighting in the assimilation pattern presented above, we show an analysis of the identification choices together with a token-by-token cue analysis for CF /æ/. Two predictions follow from the hypothesis that the weighting of relevant perceptual cues in the native language determines how non-native vowels will be assimilated. First, for a given non-native vowel, assimilation of vowel tokens should include good instances of each assimilation percept and ambiguous tokens that are not consistently labeled. Second, variability in labeling across different tokens of the same non-native vowel should be related to the underlying cues and their weighting in the native language. As shown in Table 2, CF /æ/ tokens were labeled as both ‘æ’ and ‘ɛ’ by CE listeners. Therefore, we expect that some CF /æ/ tokens will be assimilated to ‘ɛ’ almost exclusively, some to ‘æ’ almost exclusively, and some to both ‘æ’ and ‘ɛ’ (‘?’) with neither response being chosen consistently. In addition, the good ‘ɛ’ percepts should cluster in cue space close to English /ɛ/, the good ‘æ’ tokens should cluster in cue space close to English /æ/ and the inconsistent percepts should fall in an intermediate space between these two clusters.

To assess these predictions, we examined the assimilation response patterns for each of the 30 CF /æ/ tokens. We found that 20% of the CF /æ/ tokens were assimilated to ‘ɛ’ more than 90% of the time; 30% to ‘æ’ more than 90% of the time; and 17% were assimilated as both ‘ɛ’ and ‘æ’ (‘?’) with each response chosen between 40% and 60% of the time. Thus, assimilation of individual CF /æ/ tokens shows the kind of variability that we expected. To determine whether this variability is related to the weighting of spectral and duration cues in the native language, we plotted each token type (good ‘æ’, good ‘ɛ’, and ambiguous ‘?’) in a cue space defined by F1 (vertical) and duration (horizontal) as shown in Figure 5. These identification types are shown in gray; and, shown in black, are the average F1 and duration values observed in our analysis of English /ɛ/ and /æ/ productions in the same phonetic contexts. The data in Figure 5 show that, as expected, good ‘ɛ’ tokens are short and high and cluster close to English /ɛ/ while good ‘æ’ tokens tend to be long

and low and cluster close to English /æ/. Also ambiguous tokens fall between the 'ε' and 'æ' token types and are intermediate in vowel height and in duration. Thus, overall, assimilation patterns appear to be based on the relative weighting of spectral and duration cues in the listeners' native language.



Figure 5: F1 and duration values for the CF /æ/ tokens (produced in multiple consonantal contexts) perceived as English 'æ', English 'ε', or both '?'. English average vowel acoustics are shown in black.

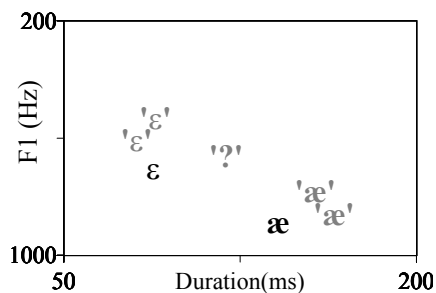


Figure 6: F1 and duration values for the CF /bæs/ perceived as English 'æ', English 'ε', or both '?'. English average vowel acoustics are shown in black.

As mentioned above, assimilation patterns are known to differ when the same non-native vowel occurs in different phonetic contexts. For this reason, we also examined the cue/response relationship in CE perception of CF /æ/ tokens produced in the same CVC context. Figure 6 shows the 3 token types observed for CE perception of 5 of 6 CE /bæs/ tokens presented in the assimilation task. The token types are shown in gray and the black ε and æ symbols show the average F1 and duration values for English /bæs/ and /bes/ tokens. Here also, good 'ε' tokens are short and high and cluster with English /ε/; good 'æ' tokens are long and low and cluster with English /æ/; and the ambiguous token falls between /ε/ and /æ/. Thus, the multiple-category assimilation observed in the CE perception of CF /æ/ occurs even when phonetic context does not vary and appears to be related to the perceptual weighting of spectral and durational cues in the listeners' native language. Similar findings are evident when cue/response relationships are examined for the other CF vowels.

4. IMPLICATIONS FOR L2 LEARNING

From the cross-language identification results, it seems clear that CE beginning learners of CF will misidentify

many CF vowel tokens so long as they rely on their English cue weighting scheme. However, not all vowels are likely to show equal degrees of misidentification. The findings reported here suggest that a useful general solution for CE learners of CF is to learn to ignore vowel duration. This may turn out to be quite difficult since duration seems to be an important cue to vowel categorization for English listeners, especially when non-native vowels show considerable spectral ambiguity [4] and since vowel duration has been suggested to be a universally preferred dimension in cross-language vowel identification studies [3]. Our findings also suggest that CE learners of CF must learn to tune into the spectral properties that specify CF vowels. Specifically, boundaries between spectrally neighboring front vowels will have to shift in order to access functional differences in CF. Clearly, those boundaries can be properly re-aligned by an overall upward shift in perceptual F1 boundaries for front vowels. However, the best way to achieve this perceptual change is unknown. Perhaps, CE listeners must redefine the F1 boundaries specifying each front CF vowel category, or perhaps, once one CF vowel category is well established, it will have an advantageous influence on the perception of its spectral neighbor. In either case, L2 perception training on CF vowels should be beneficial if it can facilitate the attunement to the relevant spectral properties and diminish the perceptual reliance on durational information.

REFERENCES

- [1] C. Best. "A direct realist view of cross-language speech perception", in W. Strange, Ed., *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, pp. 171-204, 1995.
- [2] W. Strange, R. Akahane-Yamada, R. Kubo, S. A. Trent and K. Nishi. "Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners", *Journal of the Acoustical Society of America (JASA)*, vol. 109, pp. 1691-1704, 2001.
- [3] P. Escudero, "The role of the input in the development of L1 and L2 sound contrasts: language-specific cue weighting for vowels", *Proceedings of the 25th Annual Boston University Conference on Language Development*, pp. 250-261, 2001.
- [4] J. M. Hillenbrand, M. J. Clark and R. A. Houde, "Some effects of duration on vowel recognition", *JASA*, vol. 108, pp. 3013-3022, 2000.
- [5] T. Gottfried and P. Beddor, "Perception of temporal and spectral information in French vowels", *Language and Speech*, vol. 31, pp. 57-75, 1988.
- [6] M. Picard, *An Introduction to the Comparative Phonetics of English and French in North America*, Amsterdam and Philadelphia: John Benjamins, 1987.