Article

The Mixed Effects of Phonetic Input Variability on Relative Ease of L2 Learning: Evidence from English Learners' Production of French and Spanish Stop-Rhotic Clusters

Laura Colantoni [1] and Jeffrey Steele [2,*]

[1]   Department of Spanish and Portuguese, University of Toronto, 91 Charles St W,
     Toronto, ON M5S 2C7, Canada; laura.colantoni@utoronto.ca
[2]   Department of French Studies, University of Toronto, 50 St Joseph Street,
     Toronto, ON M5S 1J4, Canada
[*]   Correspondence: jeffrey.steele@utoronto.ca; Tel.: +1-416-926-1300 (ext. 3352)

check for updates

Abstract: We examined the consequences of within-category phonetic variability in the input on non-native learners' production accuracy. Following previous empirical research on the L2 acquisition of phonetics and the lexicon, we tested the hypothesis that phonetic variability facilitates learning by analyzing English-speaking learners' production of French and Spanish word-medial stop-rhotic clusters, which differ from their English counterparts in terms of stop and rhotic voicing and manner. Crucially, for both the stops and rhotics, there are differences in within-language variability. Twenty native speakers per language and 39 L1 English-learners of French ($N = 20$) and Spanish ($N = 19$) of intermediate and advanced proficiency performed a carrier-sentence reading task. A given parameter was deemed to have been acquired when the learners' production fell within the range of attested native speaker values. An acoustic analysis of the data partially supports the facilitative effect of phonetic variability. To account for the unsupported hypotheses, we discuss a number of issues, including the difficulty of measuring variability, the need to determine the extent to which learners' perception shapes intake, and the challenge of teasing apart the effects of input variability from those of transferred L1 articulatory patterns.

Keywords: input; variability; relative difficulty; second language production; consonant clusters; stops; rhotics; phonetic parameters; voicing; manner

1. Introduction

Research on the role of input in L2 speech learning has first and foremost examined the effects of overall input quantity and quality (e.g., Flege 2009; Freed et al. 2004; Moyer 2009, 2011; Saito 2015, for phonetics and phonology). A smaller body of work has studied the effects of input variability, namely, the extent to which the latter may facilitate speech learning. As we will see in greater detail in Section 2, findings have overwhelmingly found a positive role for this variable. Laboratory studies investigating native talker variability—specifically, learners' exposure to one versus many speakers, including speakers of different varieties—have shown that such variability not only improves the discrimination of some native contrasts (the English /l/-ɹ/ contrast in particular) but also the retention of categories in long-term memory (e.g., Logan et al. 1991; Pisoni and Lively 1995).

We seek to contribute further to the study of the role of input variability in L2 speech learning by expanding on previous research in several ways via the analysis of the acquisition of French and Spanish word-medial stop-rhotic (SR) clusters (e.g., French *sucré* /sykʁe/ 'sweet', *degré* /dəgʁe/ 'degree';

Spanish: *sobra* /so<u>br</u>a/ 'excess', *sidra* /si<u>dr</u>a/ 'cider') by native speakers of English. First, whereas the majority of previous research has looked at variability's effects on the perception of phonemic contrasts, here, we seek to determine whether such effects are also observed in the production of subphonemic properties. Second, in contrast to some researchers (e.g., Logan et al. 1991; Pisoni and Lively 1995), we focus on within-category variability within a given language as well as between speakers as opposed to between-speaker variation alone. Third, we analyze the interaction of variability and proficiency by comparing intermediate versus advanced speakers. The present research thus differs from much previous research, which has looked at L2 learners at the initial stages of L2 speech learning. Finally, we make two new empirical contributions by comparing the relative ease of acquisition of consonant sequences (in contrast to individual liquids or vowels) and by investigating the acquisition of two Romance languages (as opposed to English).

The remainder of the paper is structured as follows. In the following section, we review previous empirical research on the role of acoustic variability in L2 phonetic and lexical learning. We will see that, overwhelmingly, such studies reveal the positive effect of this variable. We then present a detailed overview of the phonetic properties of French and Spanish SR clusters mentioned above, focusing particularly on patterns of variability in stop and rhotic voicing and manner. This is followed by the presentation of a set of specific hypotheses for the acquisition of these clusters by native English speakers for each of the phonetic parameters in question. These hypotheses are then investigated using data from an experimental study that tested L2 learners of intermediate and advanced proficiency on their production of these clusters via a carrier-sentence reading task. We conclude with a discussion of the importance of variability as well as other factors, including first language influence and articulatory complexity, as predictors of relative difficulty in L2 speech learning.

2. Empirical Evidence for the Role of Input Variability in L2 Acquisition

Laboratory studies have generally demonstrated that exposure to multiple talkers or larger (i.e., more variable) stimuli sets seems not only to facilitate category formation in perception (e.g., Brousseau-Lapré et al. 2013; Nishi and Kewley-Port 2007, for vowels; Logan et al. 1991; Pisoni and Lively 1995; Pruitt et al. 2006; Sadakata and McQueen 2013; Zhang et al. 2009, for consonants; Hardison 2003, for prosody) but also the retention of these categories in memory (Lively et al. 1994; Pisoni and Lively 1995) and the ability to extend training to novel contrasts/speakers (e.g., Clopper and Pisoni 2004; Nishi and Kewley-Port 2007; Pruitt et al. 2006; Sadakata and McQueen 2013). Indeed, (Brousseau-Lapré et al. 2013, p. 420) claim that "there is now a consensus that highly variable natural speech input provides the best foundation for learning in second language speech perception interventions". Some caution in accepting such consensus may be warranted given that other research has either shown no advantage or even disadvantage for more heterogeneous speech input. Iverson et al.'s (2005) training study showed no greater improvement using input with signal manipulation than with high-variability phonetic training; this suggests that it may be training in general—as opposed to stimuli speaker variability in particular—that shapes learning. Giannakopoulou et al. (2017), in a study of native Greek speakers' perception of the English /i-ɪ/ contrast, found greater advantage for training with a single, as opposed to multiple, talkers and that this benefit increased over the course of the 10 training sessions. Finally, Bohn and Bundgaard-Nielsen (2009, p. 218) demonstrated that the least intelligible vowels produced by their Danish-speaking L2 learners were the same vowels that vary the most across English (American, Southern British, and Australian) dialects. These authors propose explicitly that "an additional source of learning problems for *non-native* speakers is inherent to a learning target that is highly variable".

Mixed findings regarding the effects of input variability on L2 learning can also be observed in a series of studies that tested the effects of acoustic variability on speakers' vocabulary learning (Barcroft and Sommers 2005, 2014; Sommers and Barcroft 2007). Barcroft and Sommers (2005) found improved performance in terms of both accuracy and reaction time for learners trained with either greater between-talker variability (1 versus 3 (moderate) versus 6 speakers (high variability)) and

within-talker variability related to voice type (neutral, excited, whispered, and nasal as well as digitally edited high-pitched and elongated variants). This study also demonstrated that the beneficial effects of greater input variability may be blocked under certain conditions. Specifically, the positive effects of voice-type variability on L2 vocabulary learning were only observed, once potential between-speaker differences in intelligibility were controlled for. Sommers and Barcroft (2007) present a similar study, differing principally in the acoustic parameters targeted. A positive effect for rate of speech—but not for overall amplitude and fundamental frequency—was found. Following (Sommers et al. 1994, p. 232), the authors argue that the fact that the positive effect on vocabulary learning was limited to variability in speech rate is consistent with the phonetic-relevance hypothesis that "acoustic variability will impair spoken word identification only if the source of variability in question alters phonetically relevant properties of the speech signal". This hypothesis echoes the general explanation for the effectiveness of high-variability training on phonetic contrasts, namely, that the "experience of variation allows the formation of generalized representations that include only phonetically relevant cues and exclude irrelevant talker identity cues" (Giannakopoulou et al. 2017, p. 6). In other words, greater variability directs learners' attention to those aspects of the input that are relevant and most consistent across exemplars. Barcroft and Sommers (2014) tested the phonetic-relevance hypothesis directly, examining the effects of fundamental frequency (f0) variability. Consistent with the hypothesis, speakers of a tonal language (Zapotec) benefitted from such training, whereas speakers of a nontonal language (Spanish) did not. In summary, taken together, these three studies demonstrate the potential effects of input phonetic variability on L2 lexical learning, including that such effects may only manifest themselves when the phonetic parameter manipulated is relevant to an L1 phonemic contrast.

Thus, experimental studies on the L2 acquisition of phonetic and lexical competence demonstrate the mainly positive effects for input variability. However, such effects may be mitigated by the type of phonetic variability. As demonstrated in Barcroft and Sommers (2005, 2014) and Sommers and Barcroft (2007), in keeping with the phonetic-relevance hypothesis, input variability's effects may be restricted to L1 contrastive features.

We now turn to the French and Spanish SR clusters targeted in the present experiment, with the goal of highlighting differences in the degree of variability in stop and rhotic voicing and manner in order to be able to propose a set of variability-based hypotheses.

3. The Phonetics of French and Spanish Stop-Rhotic Clusters

Stop-rhotic clusters differ in English, French, and Spanish in at least three respects.[1] First, English speakers must master Romance stop voicing; this involves eliminating aspiration in voiceless stops and realizing phonemically voiced stops as fully voiced.[2] Beyond these general differences, voiceless and voiced stops are realized variably in both languages. In French, both voiced and particularly voiceless stops may be variably voiced throughout their realization. In Spanish, voiced stops vary in manner: in intervocalic position, they are realized most often as approximants, but they are realized as stops after nasals and in utterance-initial position.[3] Second, the English alveolar approximant [ɹ] must be replaced by the velar/uvular fricative [ɣ]/[ʁ] in French versus the alveolar tap [ɾ] in Spanish. Once again, these are not the only possible realizations in each language. In French, [ɣ]/[ʁ] may be variably realized as an approximant (e.g., Colantoni and Steele 2007a; O'Shaughnessy 1982) and may be devoiced following a voiceless stop (e.g., Léon 1992; Tranel 1987; Walker 1984, 2001). In Spanish, the rhotic may be realized as a tap or an approximant (e.g., Blecua 2001, 2008).

---

[1]　Along with the differences to be discussed immediately, these languages also contrast in the phonetic realization of the sequences. In French and Spanish, but not in English, the clusters may be broken up by an epenthetic vowel (Colantoni and Steele 2005, 2007a). This vowel is quasi-categorically present in voiceless and voiced stop-rhotic clusters in Spanish, whereas in French it only appears in voiced stop-rhotic clusters.

[2]　Learners should also acquire different durational parameters. Although in previous studies (e.g., Colantoni and Steele 2007b, 2008) we have discussed the role of duration, we will not deal with this parameter here, given the complexity of the comparison.

[3]　The voiced dental is also realized as a stop following a lateral.

In the present study, we focused on four phonetic parameters of word-medial SR clusters, namely, stop voicing and manner, and rhotic voicing and manner. The choice to restrict the focus to SR clusters in this position was doubly motivated. First, testing word-medial as opposed to word-initial clusters ensures comparable phonetic context across speakers. When reading carrier sentences in which only the target word changes, as was done in the present study, speakers—particularly those of lower proficiency—may sometimes pause before the target word for emphasis. In contrast, word-medial clusters are realized intervocalically without exception. As pauses affect the realization of voicing, one of the parameters measured here, restricting the focus to word-medial contexts was necessary. Second, as will become obvious when discussing the results, investigating four phonetic parameters in two languages creates a large, complex data set. Focusing on two prosodic contexts (word-initial and word-medial) would only increase this complexity.

Many researchers have investigated the phonetics of stops and rhotics in singletons and clusters, but there are no previous studies comparing the relative degree of variability of stops and rhotics in such clusters along these four phonetic parameters. Thus, in order to develop the specific hypotheses for the present study, we re-analyzed data from the studies reported in Colantoni and Steele (2005, 2007a, 2011). In these studies, 40 native speakers (10 each of Quebec and European French; Argentine and Chilean Spanish) were tested on their production of obstruent-liquid clusters via the same sentence-reading task used with the L2 learners in the study reported here. Stimuli were controlled for obstruent place and manner (stops and fricatives) and liquid type (laterals and rhotics).[4] For our present needs, we focused on the subset of word-medial SR clusters (French *n* = 12; Spanish *n* = 10) in Table 1.

Table 1. French and Spanish word-medial stop-rhotic (SR) clusters.

Language	Stop Voicing					
	Voiceless			Voiced		
French (*n* = 12)	*mépris*	/meˈpʁi/	'disdain'	*cobra*	/kɔˈbʁɑ/	'cobra'
	vitré	/viˈtʁe/	'glass, ADJ'	*poudrer*	/puˈdʁe/	'powder, INF'
	sucré	/suˈkʁe/	'sweet'	*degré*	/dəˈgʁe/	'degree'
	soprano	/sɔpʁaˈno/	'soprano'	*librairie*	/libʁeˈʁi/	'bookstore'
	détraquer	/detʁaˈke/	'upset, INF'	*redresser*	/ʁədʁeˈse/	'straighten up, INF'
	décrasser	/dekʁaˈse/	'clean up, INF'	*dégrader*	/degʁaˈde/	'degrade, INF'
Spanish (*n* = 10)	*lepra*	/ˈlepɾa/	'leprosy'	*cobrá*	/koˈbɾa/	'charge, IMP, 2PS'
	letra	/ˈletɾa/	'letter'	*sobra*	/ˈsobɾa/	'left over, PRES, 3PS'
	sacra	/ˈsakɾa/	'sacred'	*podré*	/poˈdɾe/	'be able, FUT, 1PS'
	lucrá	/luˈkɾa/	'speculate, IMP, 2PS'	*sidra*	/ˈsidɾa/	'cider'
				negra	/ˈnegɾa/	'black, FEM'
				lográ	/loˈgɾa/	'achieve, IMP, 2PS'

Note: INF: infinitive; 2PS: 2nd person singular; IMP: imperative; 3PS: 3rd person singular; FUT: future; 1PS: 1st person singular; FEM: feminine.

All of the target words were read in a carrier sentence (French: *Je dis____encore une fois*; Spanish: *Digo____otra vez*; 'I say____ again') three times, each time in random order. The values reported in the following sections are based on the 1236 elicited tokens (French voiceless: 333, voiced: 374; Spanish voiceless: 237, voiced: 292) involving word-medial SR clusters.

For each member of the cluster, we measured the % voicing (proportion of visible f0 over the duration of the segment) using the same procedure as in Snoeren et al. (2006). Pulses were also displayed to verify the analysis. Manner was transcribed and transcriptions were verified based on the overall acoustic characteristics of the sounds, as determined by the visual inspection of the waveform

---

[4] These variables were controlled for in order to test a series of hypotheses that concern liquid duration and the effects of voicing, place, and stress on the duration of both the liquid and the epenthetic vowel breaking up the clusters. No effect of stress on either manner realization or percentage voicing is expected.

and spectrogram. Results were measured for each of the 40 native speakers. In what follows, we report means as well as measures of dispersion (standard deviation, range).

3.1. Stop Voicing

As a measure of phonemic stop voicing, we will focus on one particular phonetic parameter, namely, laryngeal voicing as measured by the percentage of the stop's articulation during which the fundamental frequency (f0) is present in the spectrogram. We have chosen this parameter, as opposed to the widely used voice onset time (VOT), because it facilitates both within- and between-category comparisons in two respects. First, underlying voiceless stops may be fully voiced in Romance languages, and % voicing, as opposed to VOT, better captures this reality (Möbius 2004; Snoeren et al. 2006). Second, voiced stops may be variably realized as approximants: using VOT would not allow us to compare voicing across these different manner realizations.

Figure 1 presents the density plots of the distribution of % voicing for the stop of word-medial SR clusters in both varieties of French and Spanish, controlling for the phonemic voicing of the stop (voiceless versus voiced); Table 2 displays the summary statistics for this parameter.[5] Both the density plots and the summary statistics reveal a larger dispersion in the observed values for voiceless stops. Indeed, as indicated by the results of a two-sample test of variance, there is significantly greater variability in French voiceless (SD = 32) than voiced clusters (SD = 21; $F = 2.25$, $p < 0.0001$).[6] A similar pattern is found with the Spanish native speakers (voiceless: SD = 29; voiced: SD = 19; $F = 2.43$, $p < 0.0001$). In both languages, variability is greater in voiceless clusters (French 69%; Spanish 50%) than voiced ones (French 26%; Spanish 27%). These two measures taken together show that, when determining the prototypical percentage of laryngeal voicing necessary to realize the stops of SR clusters, there is greater input variability in terms of how individual speakers produce this parameter on average as well as how such mean values vary across a range of native speakers.
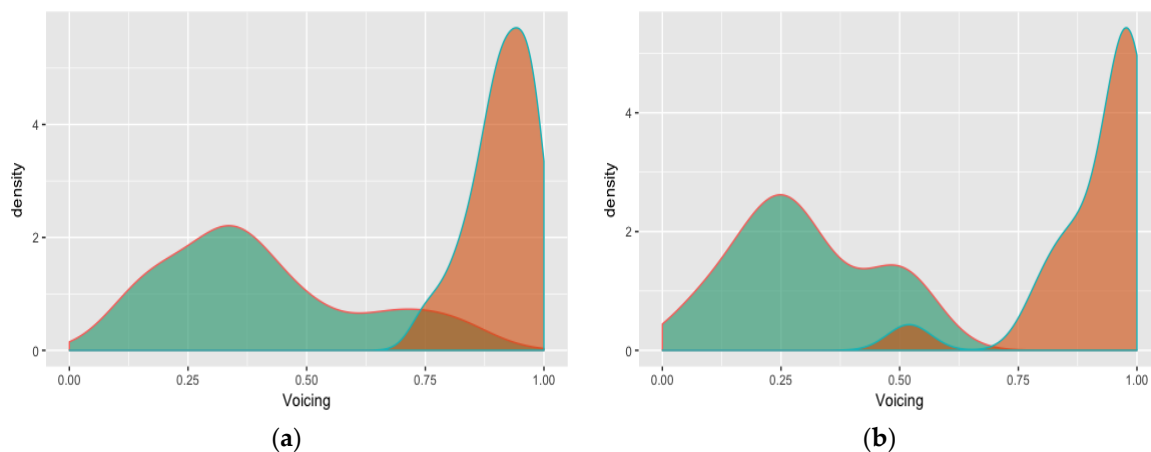

(**a**)　　　　　　　　　　　　　　　　　　　　　　　(**b**)

Figure 1. Density plots of stop voicing (%) in French (**a**) and Spanish (**b**) voiceless (green, left) and voiced (orange, right) word-medial stop-rhotic (SR) clusters.

5　See Section 4.3 "Data Preparation and Analysis" for discussion of the segmentation and measurement of the phonetic parameters of the SR clusters.

6　In order to compare the relative degree of variability for a given parameter, we used paired-sample *t*-tests of variance for quantitative variables and paired-sample *z*-tests of proportions for categorical variables following Wade et al. (2007).

Table 2. Stop % voicing: summary statistics.

	Voiceless			**Voiced**		
Language	**M**	**SD**	**Min–Max (Range)**	**M**	**SD**	**Min–Max (Range)**
French	51	32	23–92 (69)	91	21	74–100 (26)
Spanish	36	29	11–61 (50)	95	19	73–100 (27)

3.2. Stop Manner

The density plots in Figure 2 below present the variation in manner for phonemically voiceless and voiced stops in native French and Spanish; Table 3 provides the summary statistics.
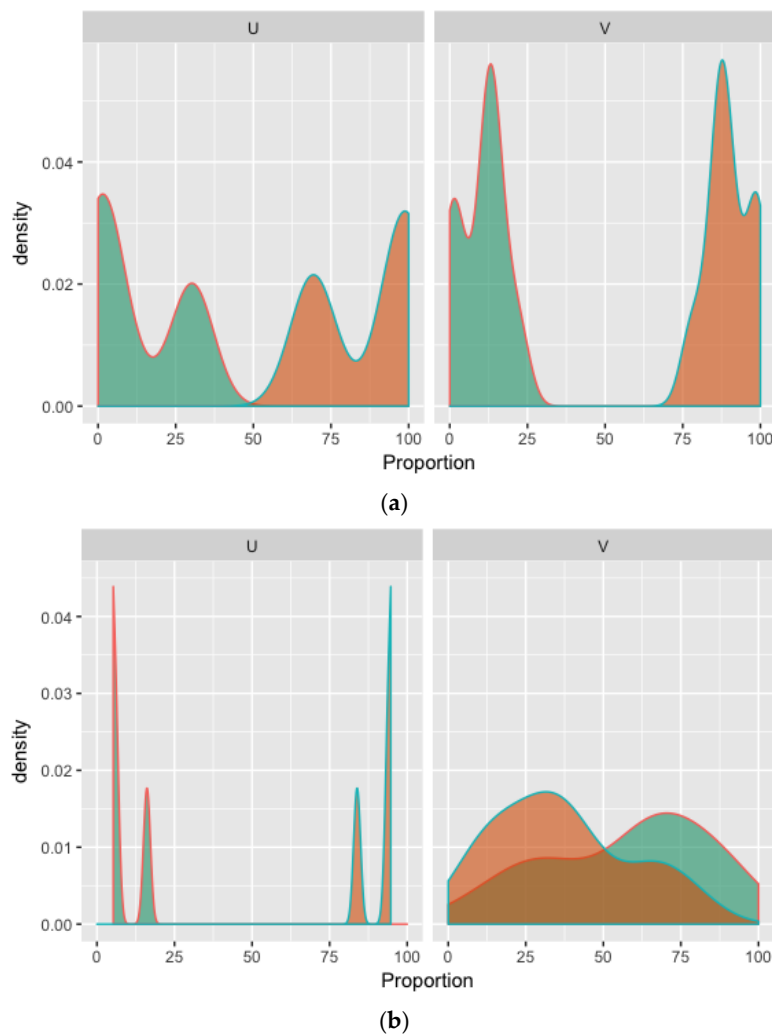

(a)

(b)

Figure 2. Density plots of stop manner (%) in French (**a**) and Spanish (**b**) voiceless (U) and voiced (V) word-medial SR clusters. Approximant (green); stop (orange).

Table 3. Stop manner: summary statistics.

	Voiceless				**Voiced**			
Language	**M**		**SD**	**Min–Max (Range)**	**M**		**SD**	**Min–Max (Range)**
	Stop	**App**			**Stop**	**App**		
French	81	19	23	50–100 (50)	86	14	12	62–100 (38)
Spanish	98	2	4	90–100 (10)	17	83	18	0–67 (67)

In French, voiceless stops are characterized by a higher degree of within-category variation (SD = 23) than voiced stops (SD = 12).[7] The relative degree of variability is also evidenced by the range of values observed; voiceless stops display more interspeaker variability (range = 50) than voiced stops (range = 38). When compared to French, Spanish seems to maintain a clear between-category distinction: voiceless stops display almost no variation (SD = 4; range = 10), whereas Spanish voiced stops display both a higher standard deviation and, particularly so, range (SD = 18; range = 67). Results of a two-sample test for proportions indicate that the proportion of Spanish stop versus approximant realizations is significantly higher than the proportion obtained for Spanish voiceless stops (z = 18.62, *p* < 0.0001).

Based on these measures of manner, within French, voiceless stops should present the least difficulty for learners. In Spanish, in contrast, it is the voiced stops that should be easier for learners to master—manner in Spanish voiceless stops should present relatively greater challenge due to the low interspeaker variability. While this latter hypothesis is strictly in keeping with the general hypothesis that L2 speech learning is facilitated by greater input variability, the degree of variation in manner in Spanish voiceless stops is so small that one could classify this parameter as invariant and thus propose that a prediction based on input variability alone is unwarranted here. We nonetheless tested the hypothesis that manner should be acquired more readily in voiced stops in order to push our general hypothesis to its limits. In the Discussion (Section 6.2), we will return to the question of what degree of variability is relevant for determining the relative difficulty of L2 speech learning.

3.3. Rhotic Voicing

Both the French voiced dorsal fricative /ʁ/ and the Spanish voiced alveolar tap /ɾ/ differ from the English voiced alveolar approximant /ɹ/. Accordingly, unlike stops where it is a matter of adjusting English phonetic parameters to target French and Spanish ones, in the case of rhotics, learners must acquire completely new articulatory patterns for French—while English has voiced fricatives (e.g., *themselves* [ð̠əmsɛlv̠z]), none are dorsal—and relatively new patterns for Spanish; the Spanish tap resembles the North American English flap allophone of intervocalic unstressed /t/ (e.g., *bottom* /bɑtəm/ → ['bɑɾəm]).

Figure 3 and Table 4 illustrate the degree of variability in the data for the first of the two rhotic parameters, rhotic % voicing.

In French (Figure 3a), significantly more variability is attested in voiceless clusters (F = 1.65, *p* < 0.0001): while this difference is minimal in terms of the standard deviation in the group mean (voiceless 34; voiced 28), the difference in the range of individual mean values is large (voiceless 81; voiced 27). In Spanish (Figure 3b), the variability in voiceless and voiced clusters is more similar. While the standard deviation is significantly greater in voiceless clusters (34 versus 26 for voiced; F = 1.73, *p* < 0.0001), the individual mean value is greater in voiced clusters (64 versus 51).

Table 4. Rhotic % voicing: summary statistics.

Language	Voiceless			Voiced		
	M	**SD**	**Min–Max (Range)**	**M**	**SD**	**Min–Max (Range)**
French	25	34	4–85 (81)	89	28	73–100 (27)
Spanish	85	34	49–100 (51)	91	26	36–100 (64)

---

[7] Results of a two-sample test for proportions indicate that the proportion of stops versus approximant realizations with French voiceless stops is not significantly higher than for French voiced ones (z = −0.55; *p* = 0.58). It is important to keep in mind that this test measures differences in proportions as opposed to differences in variability, which is reflected in the standard deviation and the range.
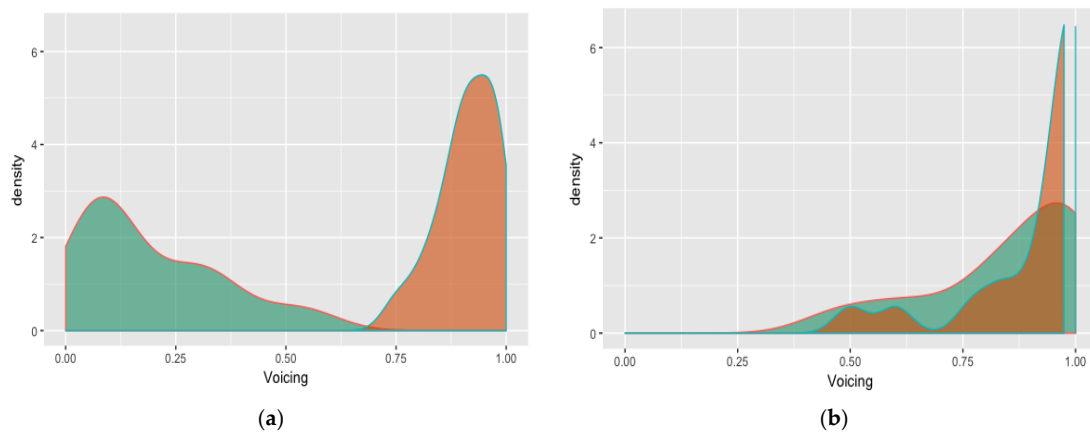
Figure 3. Density plots of rhotic voicing (%) in French (**a**) and Spanish (**b**) voiceless (green) and voiced (orange) word-medial SR clusters.

3.4. Rhotic Manner

Spanish is once again characterized by a lesser degree of variability with rhotic manner (Figure 4 and Table 5), especially with voiced clusters. Spanish /ɾ/ is realized as a tap in the majority of cases (voiceless 93%; voiced 98%) although a two-sample test of proportion reveals that the proportion of taps is significantly smaller in voiceless versus voiced clusters ($z = -2.65$, $p = 0.008$).
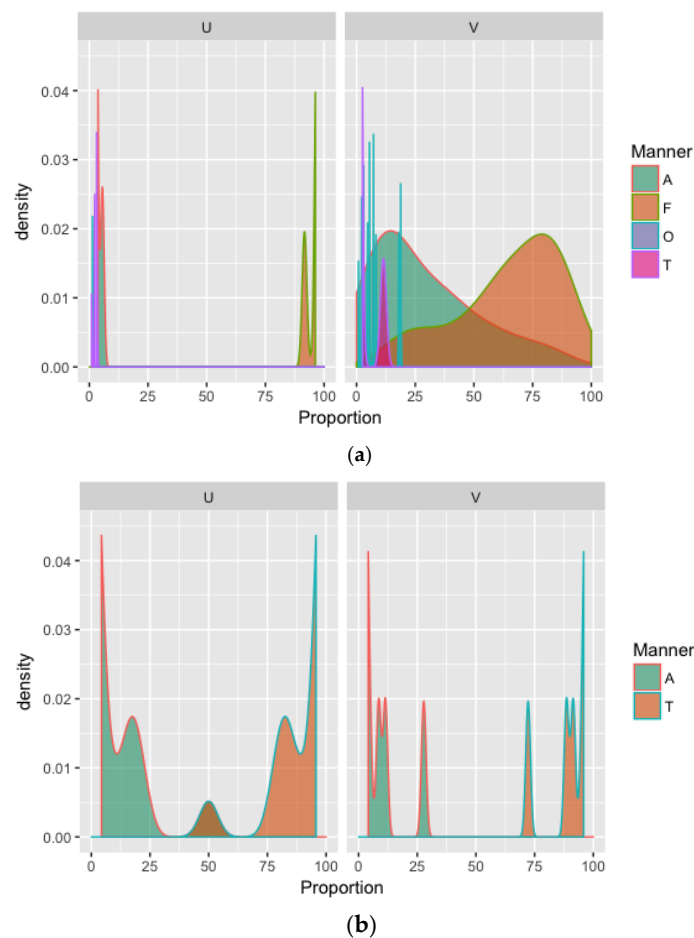

Figure 4. Density plots of rhotic manner (%) in French (**a**) and Spanish (**b**) voiceless (U; left) and voiced (V; right) word-medial SR clusters.

Table 5. Rhotic manner: summary statistics.

Language	Voiceless					Voiced				
	M			SD	Min–Max (Range)	M			SD	Min–Max (Range)
	Tap	Fric	App			Tap	Fric	App		
French		99	1	2	94–100 (6)		56	44	25	11–88 (77)
Spanish	93	6	1	10	73–100 (27)	98	2		6	73–100 (27)

Note: "SD" and "Min–Max (Range)" values are based on the percentage of fricatives in French and taps in Spanish.

In stark contrast, while the French rhotic is generally a fricative in voiceless clusters (99%), in voiced contexts, it is realized as an approximant in almost half (44%) of realizations. The difference in proportion of fricatives between voiceless and voiced contexts is significant ($z = 12.71$, $p < 0.0001$).

3.5. Summary of Variability in Stop and Rhotic % Voicing and Manner

Table 6 provides a summary of the relative within-language variability for the four phonetic parameters discussed. While the measurements presented in the preceding sections also allow for between-language comparisons, in order to test hypotheses based on differences between target languages, the characteristics of the learner groups (target language (TL) proficiency at time of testing, and quantity and quality of input encountered over the course of acquisition, among other variables) would have to be extremely similar. As we acknowledge differences in the profiles of the groups of English-speaking French and Spanish learners who participated in the present study (see Section 4.1 "Participants" below for further details), no between-language hypotheses will be proposed or tested here. However, we underline the interest and importance of doing so in future research: examining the acquisition of two different languages by learners sharing the same L1 allows for the teasing apart of the respective contribution of transfer and universal developmental effects and, as we will discuss in Section 6, both types of effects may mitigate the role of input variability on L2 speech learning.

For each combination of parameter (stop % voicing and manner; rhotic % voicing and manner) and language (within French, within Spanish), the cluster type indicated (voiceless/voiced) is the one for which there is greater variability in native speakers' production. Based on input variability alone, these are consequently the clusters for which the parameters in question are predicted to be relatively easier to acquire. Note that in the case of Spanish rhotic manner, the difference in the proportion of taps (voiceless 93%; voiced 98%), while statistically significant, is so small in absolute terms that we consider this difference not to be real.

Table 6. Summary of relatively greater within-language (French, Spanish) phonetic variability in word-medial SR clusters by stop and rhotic phonetic parameters (% voicing, manner).

Parameter	Relative Greater Variability	
	Within French	**Within Spanish**
Stop		
% Voicing	Voiceless	Voiceless
Manner	Voiceless	Voiced
Rhotic		
% Voicing	Voiceless	N/A
Manner	Voiced	N/A

On the general assumption that relatively greater within-category variability in the input has a positive effect on the acquisition of target structures, we made the following specific predictions concerning the relative difficulty of acquiring the above four phonetic parameters in voiceless versus voiced SR clusters for French and Spanish respectively:

Hypothesis 1. *In French, learners will acquire stop voicing and manner as well as rhotic % voicing more readily in voiceless clusters; rhotic manner will be easier to acquire in voiced clusters.*

Hypothesis 2. *In Spanish, acquiring stop voicing should be easier in voiceless clusters, whereas acquiring stop manner should be easier in voiced clusters. No variability-related differences are predicted for rhotic % voicing and manner.*

With these hypotheses in mind, we now turn to the experimental study designed to test them.

4. Materials and Methods

In order to test the variability-based hypotheses outlined in the previous section, intermediate and advanced English-speaking learners of French and Spanish were tested on their production of word-medial SR clusters via the same sentence-reading task used to elicit the native French and Spanish speaker data just examined. The study outlined here received approval from the University of Toronto Research Ethics Board and all participants provided written consent. In the following sections, we outline the methodological aspects of the study not already discussed with reference to the native speakers, namely, the learners' profiles, the way in which proficiency was determined via an additional production task, and further details of the data analysis.

4.1. Participants

Thirty-nine English-speaking learners participated (10 each of intermediate and advanced proficiency for French; 9 intermediate and 10 advanced Spanish learners). Given that the learners were to be tested on low-level phonetic parameters, it was important to ensure that they had been exposed to sufficient native speaker input. Indeed, in contrast to the majority of previous research on the role of input phonetic variability that has looked at the L2 acquisition of phonemic contrasts or lexical items by inexperienced learners, in order to become aware of the type of variability studied here, learners arguably require considerable experience with the target language, including with a variety of native speakers. Thus, during recruitment, learners were asked to have spent a minimum immersion period in a French- or Spanish-speaking milieu (intermediate: 3 months; advanced: 6 months). The lower range of the French intermediate speakers is due to the fact that two learners had not spent time in a French-speaking context. However, both individuals had undertaken French immersion schooling in Canada during which they would have been exposed to a wide variety of speakers over many years. In Spanish, we find quite the opposite distribution; namely, the intermediate speakers have spent, on average, more time in Spanish-speaking contexts than the advanced speakers. This difference is due to the presence of one outlier: one of the participants in the intermediate group had lived in Chile for approximately 30 years. Table 7 summarizes other relevant aspects of the learners' profiles.

Table 7. Learner profiles.

Group	Age (Years)				Immersion Experience (Months)	
	Onset of Acquisition		At Testing			
	M	Range	M	Range	M	Range
French						
Intermediate	11	5–21	33	21–52	8	0–36
Advanced	9	4–13	40	26–61	49	1–180
Spanish						
Intermediate	21	14–34	39	18–66	66	1–408
Advanced	18	11–29	30	21–35	14	3–36

Proficiency levels, which were first established prior to testing based on learners' self-evaluation, were verified via the information gathered through a conversation with each speaker and the use of a background questionnaire, as well as through a reading passage administered as part of the experiment. As is typical of many L2 phonetic production studies (e.g., Bongaerts et al. 2000; Birdsong 2008; Saito et al. 2016), accentedness scores were used as an objective measure of target language oral proficiency. Each learner was asked to read aloud the French or Spanish version of the text "The North Wind and the Sun" (Appendix A). Following the experiment, for each language, the 20/19 learner readings were interspersed with those of five native speakers of each language. The two groups of recordings were then randomized and presented to two panels (one per language) of three native speaker judges who had no training in linguistics and were unaware of the goals of the study. Following the methodology outlined in Bongaerts et al. (2000) and Birdsong (2008), the judges were asked to rate each of the readings on a scale from 1 ("heavy accent; clearly non-native") to 5 ("no foreign accent; definitely native"). The judges' ratings (Appendix B) were consistent with the proficiency level determined during subject recruitment. The average rating for the intermediate and advanced groups was noticeably different for both target languages. (French—intermediate: mean 2.1, range 1.5–2.8; advanced: mean 3.8, range 2.8–4.5; Spanish—intermediate: mean 2.3, range 1.5–3.2; advanced: mean 3.9, range 2.8–4.7).

4.2. Tasks

The learner-participants were tested individually in a quiet room. They performed three tasks: (i) the carrier sentence-reading task involving the word-medial SR stimuli in Table 1; (ii) a mirror English task included to test a set of unrelated hypotheses and thus discussed no further here; and (iii) the reading passage discussed above to determine proficiency. Stimuli for (i) were read three times, each time in random order. Out of a possible maximum of 1290 clusters (French: 12 clusters × 20 speakers × 3 rounds = 720; Spanish: 10 clusters × 19 speakers × 3 rounds = 570), once learner realizations that involved obvious misreadings or mispronunciations (e.g., metathesis or deletion; $n = 84$) were removed, a total of 1206 clusters were available for analysis (French: voiceless 347 (intermediate 175, advanced 172), voiced 351 (intermediate 175, advanced 176); Spanish voiceless 226 (intermediate 108, advanced 118), voiced 282 (intermediate 134, advanced 148)).

Between each round, the subjects were given approximately 5 minutes of break. During the first break, they completed a background questionnaire, which provided, among other things, the information summarized in Table 7. Following the three rounds, the participants read the language-appropriate version of 'The Northwind and the Sun'. Finally, the English stimuli were read once. All four rounds were recorded (44,100 Hz; 32-bit; stereo) using a Marantz CDR300 CD recorder (Marantz, Kawasaki, Japan) and unidirectional Audiotechnica AT803B lavaliere microphone (Audio-Technica U.S., Inc., Stow, OH, USA). Participants were remunerated $10 CDN.

4.3. Data Preparation and Analysis

Sound files were downsampled (22,050 Hz; 16-bit; mono) and tokens involving word-medial SR clusters were extracted and labeled. All such tokens were analyzed acoustically using Praat 4.0.41 (www.praat.org). As stated previously, voicing was measured in terms of the percentage of the stop or rhotic's articulation involving the presence of f0[8] in the spectrogram. In order to measure the duration of the stop, we took into account several parameters (see Figure 5). The onset of the stop was determined by a drop in intensity and a lowered first formant, whereas the stop offset was signaled by the presence of a burst, a rise in F1, and a rise in intensity (voiceless stops) or by the two latter parameters alone in the case of voiced stops, and, particularly, the approximant realizations. Values for percentage voicing were then averaged over the three repetitions of each

---

[8] The autocorrelation method—the default method in PRAAT for the analysis of intonation—was used.

token. Manner—stop, fricative, tap, approximant, trill, vocalization, or realizations of mixed manner (e.g., initial approximantization followed by frication or vice versa)—was transcribed and evaluated based on examination of periodicity and noisiness in the waveform and spectrogram (i.e., acoustic information was used to verify the transcription; no specific acoustic parameters were measured). Given the low proportion of vocalizations and mixed manner realizations, they were recoded as "Other" in the results presented in Tables 8–11. Results were exported to a spreadsheet and statistics were calculated with the Statistical Analysis Software (SAS).
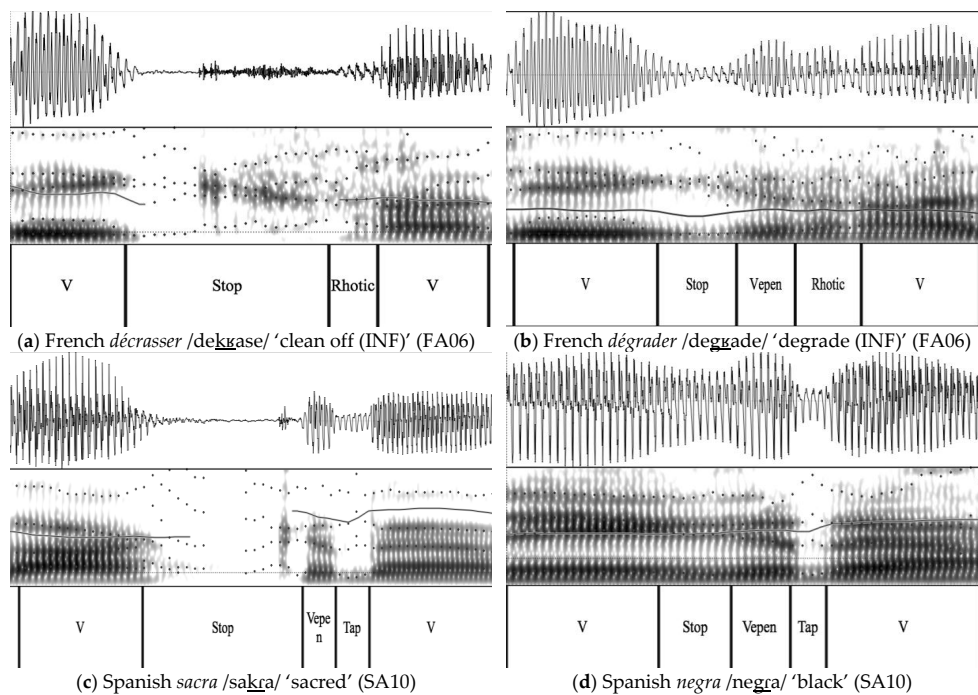

Figure 5. Sample segmentations of French (**a**) voiceless and (**b**) voiced; and Spanish (**c**) voiceless and (**d**) voiced SR clusters as realized by two advanced learners.

In order to determine whether learners had acquired a given parameter, we compared their production with that of the native speakers examined in Section 3. While cognizant of the importance of keeping in mind speaker characteristics—including the type of bilingualism—when constituting the control group in L2 acquisition studies (e.g., Grosjean 1988; Hulstijn 2012), our choice to use more French-/Spanish-dominant native speakers was triply motivated. First, the nature of the within-category voicing and manner variability in SR clusters has previously not been documented for French or Spanish with any population. It thus seemed logical to start with more monolingual, French-/Spanish-dominant speakers of the target languages. Second, all of the native speakers in both control groups had some knowledge of English and, thus, their speech was typical of the type of input to which our learners would have been exposed during their classroom learning experience. Finally, we are unaware of any literature having demonstrated L2-to-L1 influence on stop and rhotic % voicing and manner; as such, there was no empirical ground for believing that balanced bilinguals would constitute a better comparison group.

The quantitative criterion for determining whether a structure has been acquired has been the object of considerable debate in L2 acquisition research (see e.g., White 2003, pp. 77–78, for morphosyntax). In the present study, we started by using the criterion of ±2 standard deviations from the native speaker group mean. Adopting this criterion respected a central assumption concerning category formation and variability in L2 acquisition: interlanguage categories are based on archetypical values reflecting the totality of the heterogeneous input to which learners are exposed (e.g., Ellis 2003, in general; Flege 2009, p. 175, for L2 phonetics). The archetypical values for a parameter are reflected by the mean,

while the heterogeneity of the input can be measured by the standard deviation. However, for all but two of the comparisons (stop and rhotic manner in Spanish voiced SR clusters), this resulted in range minima that fell below the lowest of the attested individual native speaker means. Had the criterion of ±2 standard deviations from the native speaker group mean been kept, this would have resulted in overestimating the learners' ability, as some learners whose mean values for parameters resembled no native speaker control would nonetheless have been evaluated as having acquired those parameters. Accordingly, it was decided to use the attested native speaker group minima and maxima as the lower and upper values for a parameter to be deemed as acquired. On this criterion, learners were deemed to have acquired a phonetic parameter when their mean value for the parameter fell within the range of mean values of the native speaker controls.

5. Results

In the following sections, we will review the French then Spanish learners' production of the four phonetic parameters of the SR clusters analysed—stop % voicing and manner followed by rhotic % voicing and manner.

5.1. Stop Voicing and Manner

Tables 8–11 present the individual learner results for French and Spanish, respectively. Here and elsewhere, within the tables, learners are presented in increasing order of proficiency based on the average of the three judges' accentedness scores on 'The North Wind and the Sun' reading passage (see Appendix B for the accentedness scores). The shaded cells highlight the learner means that fall within the range of the control group's means for the parameter in question, that is, those parameters deemed to have been acquired.

Table 8. English-speaking learners' realization of stop % voicing and manner, and rhotic % voicing manner in French voiceless SR clusters. Here and elsewhere, shaded cells highlight learner means falling within the range of the individual means of the control group (i.e., that learners have acquired the parameter).

Learner	Stop			Rhotic				
	% Voicing	Manner (%)		% Voicing	Manner (%)			
		Stop	App		Tap	Fric	App	Other
FI02	16	100		93	24	76		
FI07	26	100		4		100		
FI01	20	100		88	17	11	56	16
FI10	15	100		5		100		
FI11	18	100		5		100		
FI04	75	100		28		100		
FI06	11	100		34		94	6	
FI08	13	100		7		100		
FI05	08	100		1		100		
FA06	24	100		34	19	75	6	
FI03	03	100		27		81	13	6
FA05	59	100		9		100		
FA09	40	100		4		100		
FA01	20	100		22		100		
FA10	62	100		19		100		
FA02	62	100		17		100		
FA07	28	100		15		94	6	
FA08	43	100		22		100		
FA04	49	100		3		100		
FA03	49	100		4		100		
NS Range	23–92	50–100		4–85		94–100		

App = approximant, Fric = fricative, NS = Native speaker.

Table 9. English-speaking learners' realization of stop % voicing and manner, and rhotic % voicing and manner in French voiced SR clusters.

Learner	Stop			Rhotic				
	% Voicing	Manner (%)		% Voicing	Manner (%)			
		Stop	App		Tap	Fric	App	Other
FI02	95	100		100	41	59		
FI07	30	94	6	1		100		
FI01	89	100		86	33	11	39	17
FI10	24	100		43		83	17	
FI11	57	100		62		94	6	
FI04	73	100		73		89	11	
FI06	31	94	6	65		83	17	
FI08	65	100		68		100		
FI05	67	100		56		94	6	
FA06	75	100		100	50	6	39	5
FI03	13	100		65		81	13	6
FA05	79	100		68		94	6	
FA09	56	100		97		87	13	
FA01	74	100		99		78	22	
FA10	77	100		72		89	11	
FA02	100	100		100		78	22	
FA07	100	100		95		94	6	
FA08	94	100		58		100		
FA04	82	100		72		100		
FA03	83	100		64		94	6	
NS Range	74–100	62–100		73–100		11–88		

Table 10. English-speaking learners' realization of stop % voicing and manner, and rhotic % voicing and manner in Spanish voiceless SR clusters.

Learner	Stop			Rhotic				
	% Voicing	Manner (%)		% Voicing	Manner (%)			
		Stop	App		Tap	Fric	App	Other
SI08	21	100		68	17		17	66
SI06	13	100		41	92			8
SI02	20	100		88	37	26	37	
SI10	3	100		86	87	13		
SI07	13	100		46	42			58
SI09	13	100		44	58			42
SA04	10	100		77	84	8	8	
SI04	39	100		29			17	83
SI01	2	100		59	100			
SI03	15	100		45	50	13	25	12
SA03	7	100		27	9	82	9	
SA02	52	100		67	100			
SA09	45	100		76	100			
SA10	40	100		92	58	33	9	
SA06	84	100		100	73		27	
SA07	30	100		97	100			
SA05	47	100		79	100			
SA08	37	100		51	100			
SA01	9	100		90	100			
NS Range	11–61	90–100		49–100	73–100			

Table 11. English-speaking learners' realization of stop % voicing and manner, and rhotic % voicing and manner in Spanish voiced SR clusters.

Learner	Stop			Rhotic				
	% Voicing	Manner (%)		% Voicing	Manner (%)			
		Stop	App		Tap	Fric	App	Other
SI08	33	100		70	17		17	66
SI06	96	47	53	73	100			
SI02	97	80	20	100	60		40	
SI10	55	100		100	60	20	20	
SI07	92	73	27	90	73			27
SI09	97	86	14	100	93		7	
SA04	89	86	14	96	93		7	
SI04	85	73	27	53	80	13	7	
SI01	88	60	40	72	100			
SI03	35	93	7	60	80	10	10	
SA03	84	47	53	89	80	7	7	6
SA02	100	7	93	80	100			
SA09	95	67	33	82	93			7
SA10	100	91	9	100	80		20	
SA06	99	87	13	100	60	7	33	
SA07	98	43	57	100	100			
SA05	93	88	12	100	74		13	13
SA08	100	13	87	80	100			
SA01	89	33	67	93	100			
NS Range	73–100	0–67		36–100	73–100			

Results for French % stop voicing in voiceless clusters (Table 8) indicate that three intermediate as well as nine advanced learners met the criterion for acquisition. In the case of voiced stops (Table 9), 11 learners of French (2 intermediate; 9 advanced) had acquired this parameter. As shown in Tables 8 and 9, stop manner did not present a problem for French learners, all of whom had acquired this parameter in both voiceless and voiced SR clusters. It is interesting to observe that, although manner varied in the realization of voiceless and voiced stops in many of the native speakers' production (see the density plots in Figure 2), the English-speaking learners almost categorically used stops. This may indicate that learners were using L1-based categories, which could also explain why some of the learners failed to produce target-like % voicing. It is also plausible that learners are realizing the most frequent, prototypical manner realization they encountered across native speakers; this would reinforce the use of L1 categories.

Overall, the Spanish learners performed better than the French learners. In voiceless clusters (Table 10), 13 of the learners' mean values (7 intermediate, 6 advanced) fell within the range of the control means.

With phonemically voiced stops (Table 11), only three intermediate learners produced insufficient phonetic voicing. Similar to their French counterparts, all Spanish learners were successful at acquiring stop manner in voiceless clusters (Table 10). However, the acquisition of manner in voiced clusters (Table 11), where all learners produced some percentage of approximants, proved to be extremely problematic. Recall from Figure 2 that stops are realized as approximants in 87% of the native speaker realizations of voiced SR clusters. Only 8 of the 19 learners (2 intermediate, 6 advanced) had acquired this parameter. All others produced a majority of stops.

5.2. Rhotic Voicing and Manner

Results obtained for % rhotic voicing by the L2 French speakers resemble those reported for stop voicing in terms of the learners performing relatively better as a whole with voiceless stop-rhotic clusters (Table 8) than with voiced ones (Table 9). Indeed, with voiceless clusters, 16 of the 20 learners

(7 intermediate, 9 advanced) had mean values within the range of those attested for the controls. Results for voiced clusters were much worse; only 8 of the 20 learners (3 intermediate, 5 advanced) reached criterion.

An opposite asymmetry is observed for rhotic % voicing among the L2 Spanish speakers. Here, learners were more successful at mastering rhotic % voicing in voiced (all learners) than in voiceless clusters (13 of 19; 4 intermediate, 9 advanced). The relatively higher success in voiced clusters may be attributed to voicing assimilation, given that the rhotic is preceded and followed by voiced segments.

French rhotic manner in voiced stop-rhotic clusters proved to be the most difficult of the French parameters—indeed the most difficult of any of the parameters in either language—to acquire; only 6 of 20 of the L2 learners (3 intermediate, 3 advanced) matched the controls. In voiceless clusters, 16 of 20 L2 speakers (7 intermediate, 9 advanced) had mastered this parameter.

Spanish learners were less accurate with rhotic manner in voiceless than voiced clusters, where 11 (3 intermediate, 8 advanced) versus 15 of the 19 L2 speakers (6 intermediate, 9 advanced) had acquired this parameter, respectively. In addition, Spanish learners showed a relatively clear proficiency-based pattern, with advanced speakers outperforming the intermediates with both voiceless and voiced SR clusters. This was not the case in French voiced clusters, where 3 intermediate and 3 advanced proficiency learners had target-like production values.

5.3. Summary of Results: Overall Accuracy

In presenting our general hypothesis that within-category variation facilitates the acquisition of new phonetic patterns, we proposed that ease of acquisition could be measured in terms of learners' accuracy with the different parameters in question. Consequently, for each of the target languages, in this section, we will summarize the results in terms of the total number of learners of intermediate and advanced proficiency who mastered a given parameter (overall accuracy).

Table 12 provides a summary of the total number of learners out of a maximum of 20 (French) or 19 (Spanish) who mastered each of the four parameters for both voiceless and voiced SR clusters for the two learner proficiency groups.

Table 12. Number (percentage) of English speaking-learners acquiring stop and rhotic parameters (% voicing, manner) in French (Maximum = 20) and Spanish (Maximum = 19) voiceless and voiced medial SR clusters.

Language	Voiceless				Voiced			
	Stop		Rhotic		Stop		Rhotic	
	% V	Manner	% V	Manner	% V	Manner	% V	Manner
French								
Intermediate	3	10	7	7	2	10	3	3
Advanced	9	10	9	9	9	10	5	3
All	12 (60%)	20 (100%)	16 (80%)	16 (80%)	11 (55%)	20 (100%)	8 (40%)	6 (30%)
Spanish								
Intermediate	7	9	4	3	6	2	9	6
Advanced	6	10	9	8	10	6	10	9
All	13 (68%)	19 (100%)	13 (68%)	11 (58%)	16 (84%)	8 (42%)	19 (100%)	15 (79%)

Note: % V = % Voicing.

Having now established the intermediate and advanced English-speaking learners' accuracy with each the four parameters for both target languages, we are ready to evaluate the specific hypotheses underlying the study.

5.4. Hypotheses Evaluation

The experimental study set out to test the two specific hypotheses concerning the contribution of within-category variability to learning difficulty, which we repeat below:

Hypothesis 1. *In French, learners will acquire stop voicing and manner as well as rhotic % voicing more readily in voiceless clusters; rhotic manner will be easier to acquire in voiced clusters.*

Hypothesis 2. *In Spanish, acquiring stop voicing should be easier in voiceless clusters, whereas acquiring stop manner should be easier in voiced clusters. No variability-related differences are predicted for rhotic % voicing and manner.*

Table 13 provides a summary of the evaluation of both of the hypotheses organized by target language (Hypothesis 1: French; Hypothesis 2: Spanish). Based on the summary of the results for the L2 learners presented in Table 12, each hypothesis was evaluated as follows: a given parameter was deemed to be more easily acquired for the cluster (voiceless or voiced) for which there were more learners whose mean production value fell within the range of those of the native speakers. For example, based on the fact that, for French, there was greater variability in stop % voicing in voiceless as opposed to voiced SR clusters in the native speaker controls' production, it was predicted that voicing in the former should be easier for the L2 learners to acquire. As shown in Table 13, this hypothesis was supported.

Table 13. Evaluation of the variability-related hypotheses with parameters listed predicted to be easier to acquire due to higher rates of within-category variability based on overall accuracy (intermediate and advanced learners).

	Parameter			
	Stop		**Rhotic**	
Comparison	**% Voicing**	**Manner**	**% Voicing**	**Manner**
Hypothesis 1: French				
Predicted greater ease	Voiceless	Voiceless	Voiceless	Voiced
Intermediate	$\sqrt{}$		$\sqrt{}$	X
Advanced			$\sqrt{}$	X
Hypothesis 2: Spanish				
Predicted greater ease	Voiceless	Voiced	N/A	N/A
Intermediate	$\sqrt{}$	X		
Advanced	X	X		

Note: A checkmark "$\sqrt{}$" indicates that the prediction was supported; "X" indicates that the prediction was rejected; an empty cell indicates that the data were inconclusive (i.e., equal number of learners mastering the parameter in both voiceless and voiced clusters). The Spanish rhotic parameters for which no significant difference in variability existed, and thus for which no difference in acquisitional difficulty was predicted, are indicated with "N/A".

Of the predictions in Hypothesis 1 made for French, one was supported by both the intermediate and advanced groups' performance (rhotic % voicing), another was refuted by both groups' results (rhotic % manner), a third was supported only by the intermediate learners' data (stop % voicing), and, for the final prediction concerning the greater ease of acquiring stop manner in voiceless SR clusters, the results were inconclusive: as highlighted earlier, all learners mastered this parameter in both voiceless and voiced clusters. In summary, for French, when the results for each of the proficiency-based groups are considered separately, the data relevant to evaluating the general hypothesis that within-category input variability leads to greater acquisitional ease more often support or are inconclusive (three cases each) than refute the specific predictions (two cases). In contrast, of the four predictions made in Hypothesis 2 for Spanish, three out of four comparisons refute the greater ease of acquiring stop % voicing in voiceless clusters (advanced learners) and stop manner

in voiced clusters (both proficiency levels). When the predictions made for both languages are considered together, there is slightly less support for the hypothesized positive correlation between within-category variability and ease of acquisition (four of nine conclusive comparisons).

In summary, based on both the intermediate and advanced learners' accuracy, the majority of the predictions concerning Spanish were refuted, as were the predictions regarding the acquisition of rhotic manner in French. In contrast, the data support or are consistent with the variability-based predictions that French stop and rhotic voicing should be more readily acquired in voiceless clusters. In the final section, we review those cases where our general input-variability-based hypothesis was not supported, discussing factors such as L1-based influence, universal articulatory constraints, and target language proficiency as well as other challenges inherent in integrating the role of variability when modeling L2 speech learning.

6. Discussion

Overall, the findings of the present study differ from the majority of previous research that has investigated the role of input phonetic variability in ease of L2 learning. Whereas this body of research has overwhelmingly found a positive effect for variability, the evidence presented here from the study of intermediate and advanced English-speaking learners' production of French and Spanish SR clusters is very much mixed. In an attempt to explain such differences, we begin by discussing the ways in which the current data set differs from those of previous studies as concerns target language proficiency and learners' awareness of the variability under investigation. Then, we propose that L1-based transfer and universal articulatory constraints must be taken into account, at least when making predictions for production. We also discuss the need to consider further some of the assumptions made here regarding input variability, including the types of variability learners are likely to encounter in real learning situations, their ability to parse the full range of variability in order that it become intake, as well as the potential need to distinguish between variability involving discrete variables with few variants versus continuous variables with a much higher degree of variability characteristic of the phonetic parameters that were investigated here.

6.1. Differences between the Present and Previous Research: Target Language Proficiency and Learners' Awareness of Variability

In the Introduction, we highlighted that the design of the present study differs from that of much, if not most, previous research on input phonetic variability's effect on ease of L2 learning in several ways. We discuss here those differences that might help to explain the lesser support for the general hypothesis found in the present study.

As discussed in Section 2, the effect of input variability has been studied most often via laboratory perceptual training studies involving participants with little to no experience with the target language. Accordingly, such studies provide insights first and foremost into the effect of this variable on phonemic categorization at the very earliest stages of learning. In contrast, the present study tested the same general hypothesis using production data from learners with considerable learning experience both in terms of the number of years of study and, with two exceptions, having a minimum of 3 (intermediates) or 6 months (advanced learners) of French-/Spanish-language immersion. The lack of support for many of the variability-based hypotheses found in the present study parallels the findings of Bohn and Bundgaard-Nielsen (2009). Recall that these researchers found that the least intelligible vowels produced by their Danish-speaking L2 learners were the same vowels that vary the most across English (American, Southern British, and Australian) dialects. The learners in this latter study also had formal target language learning experience (5–8 years in Denmark) although, unlike our learners, had little to no immersion experience. It may be that the effects of input variability on ease of learning are strongest at the beginning of L2 acquisition when one of the main learning objectives is target-like category formation. This hypothesis is supported by the current study's data in that three of the four

parameters that provided support for a positive role for acoustic input variability involved the less experienced intermediate learners' production.

A second important way in which the present study differs from most previous research is that, in laboratory perception studies employing high variability phonetic training, learners' attention is directed towards the contrast of interest via the task. For example, the Japanese-speaking learners of English in Logan et al. (1991) were trained using a minimal pair /l-ɹ/ identification task involving explicit feedback. In contrast, it is unclear how the reading task in the present study could have heightened learners' awareness of the type of phonetic variability involved, which is arguably below the level of conscious awareness. Given the important role attributed to awareness in L2 learning (e.g., Robinson et al. 2011), this task-based difference may be relevant. We will discuss further limitations on L2 learners' analysis of the input, including the role of awareness in Section 6.3.

6.2. Determining the Relative Importance of Input Variability versus Transfer and Articulatory Constraints when Predicting Relative Ease of Acquisition

When formulating the specific hypotheses tested in the present study, the sole factor considered was the relative degree of acoustic input variability. In the case of L2 acquisition, such variability is parsed using existing L1-influenced categories. Production too is influenced by a learner's L1 as well as by universal articulatory constraints. As such, hypotheses must consider both of these factors. Teasing the role of variability apart from transfer and articulatory complexity is not a straightforward task. In some cases, they make opposite predictions. For example, while stop voicing is more variable in French voiceless than voiced SR clusters, the learners' L1 and the target language categories are more similar in voiceless contexts. The English-speaking learners' relatively greater success with stop and rhotic % voicing in voiceless contexts—a result that is in keeping with our input-variability-based hypothesis—may indeed be related to transfer. In the case of stops in both languages, learners seem to be using their L1 voiced stop categories, which are insufficiently voiced. In the case of French rhotics, the English-speaking participants may be simply transferring their native articulatory patterns, which also involve devoicing of the following liquid.

Their accuracy with rhotic % voicing may also be related to the fact that aerodynamic constraints favor voicing assimilation of the rhotic, leading naturally to target-like low levels of voicing in /ʁ/.[9] The use of L1 articulatory patterns allows for rapid accuracy, in spite of the challenge present in analyzing the target language variability. The same can be said about the challenge of considering articulatory complexity. French L2 speakers were less successful at acquiring the rhotic parameters in voiced than in voiceless clusters; this was particularly the case of rhotic manner, which was only acquired by six speakers in the former context. This could be attributed to the fact that realizing a voiced fricative is problematic from an aerodynamic point of view: whereas fricatives require an open glottis for sufficient airflow, this glottal configuration disfavors voicing. Finally, variability in L2 production does not always mirror input variability. This is illustrated with the variation in rhotic manner observed with the intermediate speakers of Spanish. As indicated in Tables 10 and 11, these learners produced a high proportion of rhotics that fell within the category 'Other', which included retroflexed rhotics not attested in Spanish that clearly resulted from transfer from English. As such, even if input variability increases the range of possible targets, complicating the analysis of categories and thus the process of category formation, input cannot be the sole source of these differences between L2 and target variability.

---

[9] The parallel failure to produce fully voiced stops may be due to durational differences. Normally, L2 learners are less fluent than native speakers (e.g., Munro and Derwing 1998; Towell et al. 1996). As such, L2 learners tend to produce longer segments and greater duration impedes the maintenance of voicing.

6.3. Considering Learners' Perception and Analysis of the Input

Numerous studies have shown that L2 learners fail to perceive some consonantal and vocalic contrasts (e.g., Cebrian 2006; Flege et al. 1996; Flege and MacKay 2004; Guion et al. 2000) as well as consonant sequences (e.g., Kabak 2003; Matthews and Brown 2004) in the same way as native speakers. Some of these difficulties may be overcome with experience and training, whereas others present greater challenge. Given the existing evidence that L2 learners sometimes fail to perceive contrasts in the L2 that are absent in their L1, it is possible that they may fail to notice variation in the target language. This is supported by Strange's (2009) proposal that, when task complexity increases, L2 learners use a primarily phonological level of processing. This would preclude learners—at least those of lower levels of proficiency and having less experience with the target language—from being able to parse the range of phonetic variation attested in real-world communicative situations. Failure to notice the type of subphonemic phonetic variation studied here is also in keeping with the phonetic-relevance hypothesis discussed in Barcroft and Sommers' (2005, 2014; Sommers and Barcroft 2007) studies that proposes that variability is relevant only when it targets contrastive parameters. If this is the case, the types of analyses of variation that were undertaken here in order to formulate the specific hypotheses tested would simply not be possible for many learners, even those having considerable target language experience. We propose that selective, non-native-like perception offers a partial explanation for the failure of learners of Spanish to realize underlying stops as approximants in voiced clusters: not only is approximantization of underlying stops not a feature of their L1, but transferred English grapheme-phoneme mappings would only enhance the probability of learners adopting a stop analysis. It is also possible that, even if such variation is noticed, non-native speakers may interpret it differently from native speakers. Indeed, it may be the case that, particularly at initial stages, learners interpret allophonic variation in the target language as phonemic and that, only with large quantities of input over longer periods of time, are target-like analyses possible. This interpretation may be reinforced by the existence of some overlap in the phonetic realization of the phonemic categories. For example, in French, the upper range of stop % voicing in phonemically voiceless stops overlaps with that of voiced stops (23–92% versus 74–100%; Table 2).

In a related vein, in the present study, variability in the target language was said to exist whenever statistically significant differences existed in the mean value and/or range of values for a given phonetic parameter between voiceless and voiced clusters. It may be important to nuance this operationalization. It may be the case that, even when statistically significant, differences are too small to be perceptible. Recall that, when formulating the specific hypothesis concerning Spanish stop manner, we highlighted that while greater variability was attested in voiced stops, the degree of variability in their voiceless counterparts was so limited as to be able to characterize voiceless stop manner as invariant. In such cases, it is arguably the case that using between-category variability (here, voiceless versus voiced) is not motivated.

It may also be the case that L1-shaped cue weighting may result in learners being more sensitive to certain types of variability and, consequently, relatively insensitive to others. For example, Escudero (2000) demonstrated that L2 learners may focus on the primary L1 phonetic cues to a phonological contrast even when such a cue is not used by native speakers of the target language in production. In a parallel fashion, it may be the case that learners fail to perceive phonetic variability, at least at earlier stages of acquisition, when L1 realizations of the same phonological category do not vary along the same phonetic parameter(s). In such cases, input variability could not be used to predict acquisitional difficulty, as it would not constitute part of learners' intake. In future work, it will be important to define variability not only in terms of production measures—including whether there are thresholds below or above which variability is too limited to be relevant to predicting relative difficulty—but also as determined by learners' sensitivity to existing differences via perception tests. Finally, it could also be assumed that, even when learners are sensitive to these differences and are able to establish native-like long-term phonetic representations, they lack the articulatory control necessary to realize the patterns attested.

6.4. What Is Learners' Input?

Perhaps the greatest challenge to predicting the effects of variability is determining the actual input to which learners are exposed over the course of acquisition. For both French and Spanish, the specific predictions formulated here were based on the production of 20 native speakers of two different varieties. One might question whether it is reasonable to assume that all of the L2 learners in the present study, even with the requirement of a minimum of 3 or 6 months of immersion, would have encountered such a degree of variability. For some learners, particularly the most proficient, this is likely. For other learners, the range of variability encountered might be much less than that measured in the native speaker controls in the present study. Furthermore, the extent to which variability predicts ease of acquisition might depend upon the point at which it is encountered. As proposed in Section 6.1, if learners' sensitivity to variation is indeed related to phonetic properties of the L1, they might be less sensitive at earlier stages of learning. However, it is also possible that the instantiation of phonetic categories in long-term memory over the course of acquisition might desensitize L2 learners to variability if it is not encountered early enough. This would be true of those classroom learners having little exposure to speakers other than their (non-native) instructors during the first years of learning. In summary, it may be necessary not only to know what degree of variability existed in a learner's input, but also at which point along the acquisitional path it was encountered. Determining the degree of input variability at various stages of learning may be, unfortunately, impossible (see (Flege 2009) for further discussion).

6.5. Types of Variability

In future research, it will be necessary to distinguish between different types of variability and the relatively different challenges they may pose for learners. Determining the role that input variability plays is complicated by the fact that the label 'within-category' variation includes inter- and intraspeaker variation in the realization of a given category. For example, a given phoneme such as a French voiceless stop may be realized categorically as a stop by some speakers or variably as a stop or approximant by others. Another phoneme, like the French rhotic, may show a higher degree of within-category variation, varying in manner among all speakers. These two types of within-category variation should pose different problems for L2 learners. The first type should be easier for learners to manage, as they have a wider range of options for realizing a target structure, all of which constitute target-like production. The second type may be more difficult to perceive and produce. In addition to the range of within-category variation, languages vary in how sharp the contrasts between categories are, and native speakers vary on how the perceive and produce these contrasts (Perkell et al. 2006). Although not analyzed here, it is expected that sharper phonemic contrasts will pose fewer difficulties to L2 learners than cases where there is some degree of overlap between members of the categories (see Wade et al. 2007).

6.6. Measuring Acquisition and Relative Ease of Learning

We wish to conclude our discussion by exploring further the criterion used to determine whether a given structure has been acquired. In the present study, acquisition was equated with learner values falling in the range of the control means for a given phonetic parameter. Other criteria have been proposed and used in the past, as was highlighted in Section 4.3. First, we could have simply used the control group's mean as the point of reference. However, while using means would have been relatively meaningful for some of the parameters, this would not have been the case for others. For example, stop or rhotic manner in French and voiceless stop and rhotic manner in Spanish are relatively less variable than the other parameters that were explored here. While we could argue that learners may be able to calculate and use means in such cases, means are not representatives of the native speakers' performance with the other highly variable parameters. Indeed, very few native speakers have values close to the group mean; the median may indeed be more representative of typical possible values in

such cases. Moreover, in the case of highly variable parameters, it is not clear that acquiring the mean equals having acquired the target category, which is intrinsically variable. Second, as we have done in past research (Colantoni and Steele 2006), we could have undertaken a speaker-by-speaker analysis, comparing each of the L2 learners against each of the native speakers for all of the parameters. In this case, acquisition would be deemed to have occurred when a given learner matches any native speaker control for all parameters. While a useful method for evaluating native-like ultimate attainment, in the study of variability, using a speaker-to-speaker comparison would go against the idea that learners use the aggregate input in forming their categories. If this is true, and if this input is more varied than for L1 learners in terms of the range of native and non-native speakers who serve as models, it is arguably unwise to compare learner production with individual native grammars.

7. Conclusions

Explaining the great degree of variation in L2 production has been, and will continue to be, a central goal in developing theories of non-native acquisition. In the present work, we have sought to contribute to this line of research by investigating the degree to which within-category target language variability predicts the difficulty of acquisition of four phonetic parameters in French and Spanish stop-rhotic clusters. The general hypothesis that a greater degree of variability in the input results in more accurate acquisition was partially supported, in that it predicted a subset of the specific hypotheses formulated. In those cases where hypotheses were not supported, such as the acquisition of French rhotic manner or Spanish stop manner, we highlighted areas that require further consideration including (i) the difficulty in measuring the variability in the actual input to which a given learner is exposed; (ii) the need to determine the extent to which learners' perception may lead to intake which makes them less sensitive to variability including how this might be affected by target language experience and proficiency; and (iii) teasing the effects of input variability from those of transferred L1 articulatory patterns and universal production constraints.

In future research, it will be necessary to build on the findings here. This will include expanding the target structures to take into account prosodic phenomena such as stress and intonation. In particular, it will be of interest to test learners on sets of structures for which a clearly defined hierarchy of variability can be established, so as to be able to further test the general hypothesis proposed here. It would also be of interest to conduct studies involving learners with a range of target language experience in order to test directly the hypothesis that the influence of input variability may wane with time as learners move from primarily categorizing to producing the target language. On a methodological level, we have highlighted the need to couple production studies on the acquisition of variability with perception studies that will allow us to determine to what extent L2 hearers are sensitive to low-level phonetic variation and, thus, refine the general proposal that variability increases the learning challenge. In conclusion, pushing forward our understanding of the role of variability in L2 learning will require insights from both perception and production formalized in a multifactor model that takes into account all of the aspects discussed here.

Appendix A. French and Spanish Texts (Proficiency Measure)

French: "La bise et le soleil"

La bise et le soleil se disputaient, chacun assurant qu'il était le plus fort. Quand ils ont vu un voyageur qui s'avançait, enveloppé dans son manteau, ils sont tombés d'accord que celui qui arriverait le premier à le lui faire ôter serait reconnu comme le plus fort. Alors, la bise s'est mise à souffler de toutes ses forces mais plus elle soufflait, plus le voyageur serrait son manteau autour de lui. Finalement, elle renonça à le lui faire ôter.

Alors, le soleil commença à briller et au bout d'un moment le voyageur, réchauffé, ôta son manteau. Ainsi, la bise dut reconnaître que le soleil était le plus fort.

Spanish: "El viento del norte y el sol"

El viento del norte y el sol discutían acerca de cuál de los dos sería el más fuerte, cuando, de repente, pasó un viajero envuelto en una amplia capa. Al verlo, convinieron en que el primero que consiguiera quitarle la capa sería el más fuerte. El viento del norte comenzó a soplar con mucha furia, pero, cuanto más soplaba, más se aferraba el viajante a su capa, hasta que el viento norte desistió. El sol brilló entonces con todo su esplendor, e inmediatamente, el viajante arrojó su capa. Así, el viento norte tuvo que reconocer la superioridad del sol.

Appendix B. Native Speaker Judges' Evaluation of French and Spanish Learners' Proficiency Level Based on Reading Passage

Table A1. French learners.

	Intermediate					**Advanced**			
	Judge			**Average**		**Judge**			**Average**
	1	**2**	**3**			**1**	**2**	**3**	
FI01	2	2	1	1.7	FA01	4	4	2.5	3.5
FI02	1.5	2	1	1.5	FA02	4.5	3.5	4	4
FI03	3.5	3	2	2.8	FA03	5	4.5	4	4.5
FI04	2	2.5	2	2.2	FA04	4	4.5	4.5	4.3
FI05	3	3	1.5	2.5	FA05	3	3.5	3.5	3.3
FI06	2	2.5	2	2.2	FA06	3	3.5	2	2.8
FI07	1	2.5	1	1.5	FA07	4.5	4	4	4.2
FI08	2	2.5	2.5	2.3	FA08	4	4	4.5	4.2
FI10	2	2.5	1.5	2	FA09	4	4	2	3.3
FI11	2	2.5	1.5	2	FA10	4.5	3.5	3.5	3.8
Average				2.1	Average				3.8

Table A2. Spanish learners.

	Intermediate					**Advanced**			
	Judge			**Average**		**Judge**			**Average**
	1	**2**	**3**			**1**	**2**	**3**	
SI01	3	3.5	3	3.2	SA01	4.5	5	4.5	4.7
SI02	2	2	1.5	1.8	SA02	3.5	4	3.5	3.7
SI03	3.5	3	3	3.2	SA03	3	4	3.5	3.5
SI04	2.5	4	2.5	3	SA04	2.5	3	3	2.8
SI06	1.5	1.5	2	1.7	SA05	4	4	4.5	4.2
SI07	2.5	2	2	2.2	SA06	4	4	4	4
SI08	1	2	1.5	1.5	SA07	4	4	4	4
SI09	1	3	2.5	2.2	SA08	4.5	4	4.5	4.3
SI10	1.5	2	2	1.8	SA09	3.5	4	4	3.8
					SA10	3.5	4	4	3.8
Average				2.3	Average				3.9

References

Barcroft, Joe, and Mitchell Sommers. 2005. Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition* 27: 387–414. [CrossRef]

Barcroft, Joe, and Mitchell Sommers. 2014. Effects of variability in fundamental frequency on L2 vocabulary learning. *Studies in Second Language Acquisition* 36: 423–49. [CrossRef]

Birdsong, David. 2008. Nativelike pronunciation among later learners of French as a second language. In *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Edited by Ocke-Schwen Bohn and Murray J. Munro. Amsterdam and Philadelphia: John Benjamins, pp. 99–116.

Blecua, Beatriz. 2001. Las vibrantes del español: Manifestaciones acústicas y procesos fonéticos. Ph.D. thesis, Universidad Autónoma de Barcelona, Barcelona, Spain.

Blecua, Beatriz. 2008. Los sonidos vibrantes: Aspectos comunes y variación. *Language Design* 1: 23–30.

Bohn, Ocke-Schwen, and Rikke Bundgaard-Nielsen. 2009. Second language speech learning and diverse inputs. In *Input Matters in SLA*. Edited by Thorsten Piske and Martha Young-Scholten. Bristol, Buffalo and Toronto: Multilingual Matters, pp. 207–18.

Bongaerts, Theo, Susan Mennen, and Frans van der Slik. 2000. Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica* 54: 298–308. [CrossRef]

Brousseau-Lapré, Françoise, Susan Rvachew, Meghan Claynards, and Daniel Dickson. 2013. Stimulus variability and perceptual learning of nonnative vowel categories. *Applied Psycholinguistics* 34: 419–41. [CrossRef]

Cebrian, Juli. 2006. Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics* 34: 372–87. [CrossRef]

Clopper, Cynthia, and David Pisoni. 2004. Effects of talker variability on perceptual learning of dialects. *Language and Speech* 47: 207–39. [CrossRef] [PubMed]

Colantoni, Laura, and Jeffrey Steele. 2005. Phonetically-driven epenthesis asymmetries in French and Spanish obstruent-liquid clusters. In *Experimental and Theoretical Approaches to Romance Linguistics*. Edited by Randall Gess and Edward J. Rubin. Amsterdam and Philadelphia: John Benjamins, pp. 77–96.

Colantoni, Laura, and Jeffrey Steele. 2006. Native-like attainment in the L2 acquisition of Spanish stop-liquid clusters. In *Selected Proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages, Minneapolis, October 2004*. Edited by Carol A. Klee and Timothy L. Face. Somerville: Cascadilla Press, pp. 59–73.

Colantoni, Laura, and Jeffrey Steele. 2007a. Voicing-dependent cluster simplification asymmetries in Spanish and French. In *Segmental and Prosodic Issues in Romance Phonology*. Edited by Pilar Prieto, Joan Mascaró and Maria-Josep Solé. Amsterdam and Philadelphia: John Benjamins, pp. 109–29.

Colantoni, Laura, and Jeffrey Steele. 2007b. Acquiring /R/ in context. *Studies in Second Language Acquisition* 29: 381–406. [CrossRef]

Colantoni, Laura, and Jeffrey Steele. 2008. Integrating articulatory constraints into models of second language phonological acquisition. *Applied Psycholinguistics* 29: 489–534. [CrossRef]

Colantoni, Laura, and Jeffrey Steele. 2011. Synchronic evidence of a diachronic change: Voicing and duration in French and Spanish stop-liquid clusters. *Canadian Journal of Linguistics* 56: 147–77.

Ellis, Nick. 2003. Constructions, chunking, and connectionism: The emergence of second language structure. In *The Handbook of Second Language Acquisition*. Edited by Catherine J. Doughty and Michael H. Long. Malden: Blackwell Publishing, pp. 63–103.

Escudero, Paola. 2000. Developmental Patterns in the Adult L2 Acquisition of New Contrasts: The Acoustic Cue Weighting in the Perception of Scottish Tense/Lax Vowels in Spanish Speakers. Master's thesis, University of Edinburgh, Edinburgh, UK.

Flege, James. 2009. Give input a chance! In *Input Matters in SLA*. Edited by Thorsten Piske and Martha Young-Scholten. Bristol, Buffalo and Toronto: Multilingual Matters, pp. 175–90.

Flege, James, and Ian MacKay. 2004. Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26: 1–34. [CrossRef]

Flege, James, Anne Marie Schmidt, and Glenn Wharton. 1996. Age of learning affects rate-dependent processing of stops in a second language. *Phonetica* 53: 143–61. [CrossRef] [PubMed]

Freed, Barbara F., Norman Segalowitz, and Dan Dewey. 2004. Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition* 26: 275–301. [CrossRef]

Giannakopoulou, Anastasia, Helen Brown, Meghan Claynards, and Elizabeth Wonnacott. 2017. High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ* 5: e3209. [CrossRef] [PubMed]

Grosjean, François. 1988. Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* 1: 131–49. [CrossRef]

Guion, Susan, James Flege, Reiko Akahane-Yamada, and Jesica C. Pruit. 2000. An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America* 107: 2711–24. [CrossRef] [PubMed]

Hardison, Debra. 2003. Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics* 24: 495–522. [CrossRef]

Hulstijn, Jahn. 2012. The construct of language proficiency in the study of bilingualism from a cognitive perspective. *Bilingualism: Language and Cognition* 15: 422–33. [CrossRef]

Iverson, Paul, Valerie Hazan, and Kerry Bannister. 2005. Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America* 118: 3267–78. [CrossRef] [PubMed]

Kabak, Baris. 2003. The Perceptual Processing of Second Language Consonant Clusters. Ph.D. thesis, University of Delaware, Delaware, MD, USA.

Léon, Pierre. 1992. *Phonétisme et Prononciations du Français*. Paris: Nathan.

Lively, Scott, David Pisoni, Reiko Yamada, Yoh'ichi Tohkura, and Takashi Yamada. 1994. Training Japanese listeners to identify English /r/-/l/. III. Long term retention of new phonetic categories. *Journal of the Acoustical Society of America* 96: 2076–87. [CrossRef] [PubMed]

Logan, John, Scott Lively, and David Pisoni. 1991. Training Japanese listeners to identify English /r/-/l/. A first report. *Journal of the Acoustic Society of America* 89: 874–86. [CrossRef]

Matthews, John, and Cynthia Brown. 2004. When intake exceeds input: Language specific perceptual illusions induced by L1 prosodic constraints. *International Journal of Bilingualism* 8: 5–27. [CrossRef]

Möbius, Bernd. 2004. Corpus-based investigations on the phonetics of consonant voicing. *Folia Linguistica* 38: 5–26.

Moyer, Alene. 2009. Input as a critical means to an end: Quantity and quality of experience in L2 phonological attainment. In *Input Matters in SLA*. Edited by Thorsten Piske and Martha Young-Scholten. Bristol, Buffalo and Toronto: Multilingual Matters, pp. 159–74.

Moyer, Alene. 2011. An investigation of experience in L2 phonology: Does quantity matter more than quality? *The Canadian Modern Language Review* 67: 191–216. [CrossRef]

Munro, Murray, and Tracey Derwing. 1998. The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning* 48: 159–82. [CrossRef]

Nishi, Kanae, and Diane Kewley-Port. 2007. Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research* 50: 1496–509. [CrossRef]

O'Shaughnessy, Douglas. 1982. A study of French spectral patterns for synthesis. *Journal of Phonetics* 10: 377–99.

Perkell, Joseph S., Frank H. Guenther, Harlan Lane, Nicole Marrone, Melanie L. Matthies, Ellen Stockmann, Mark Tiede, and Majid Zandipour. 2006. Production and perception of phoneme contrasts covary across speakers. In *Speech Production: Models, Phonetic Processes and Techniques*. Edited by Jonathan Harrington and Marija Tabain. New York: Psychology Press, pp. 69–84.

Pisoni, David, and Scott Lively. 1995. Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Timonium: York Press, pp. 433–59.

Pruitt, John, James Jenkins, and Winifred Strange. 2006. Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *Journal of the Acoustical Society of America* 119: 1684–96. [CrossRef] [PubMed]

Robinson, Peter, Alison Mackey, Susan M. Gass, and Richard Schmidt. 2011. Attention and awareness in second language acquisition. In *The Routledge Handbook of Second Language Acquisition*. Edited by Susan M. Gass and Alison Mackey. London and New York: Routledge, pp. 247–67.

Sadakata, Makiko, and James McQueen. 2013. High stimulus variability in non-native speech learning supports formation of abstract categories: Evidence from Japanese geminates. *Journal of Acoustical Society of America* 134: 1324–35. [CrossRef] [PubMed]

Saito, Kazuya. 2015. Experience effects on the development of late second language learners' oral proficiency. *Language Learning* 65: 563–95. [CrossRef]

Saito, Kazuya, Pavel Trofimovich, and Talia Isaacs. 2016. Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics* 37: 217–40. [CrossRef]

Snoeren, Natalie, Pierre Hallé, and Juan Segui. 2006. A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics* 34: 241–68. [CrossRef]

Sommers, Mitchell, and Joe Barcroft. 2007. An integrated account of the effects of acoustic variability in first and second language acquisition: Evidence from amplitude, fundamental frequency, and speaking rate variability. *Applied Psycholinguistics* 28: 231–49. [CrossRef]

Sommers, Mitchell, Lynne Nygaard, and David Pisoni. 1994. Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America* 96: 1314–24. [CrossRef] [PubMed]

Strange, Winifred. 2009. Automatic Selective Perception (ASP) of L1 and L2 speech: A working model. Paper presented at 2nd Acoustical Society of America (ASA) Special Workshop on Speech Cross-Language Speech Perception and Variations in Linguistic Experience, Portland, OR, USA, May 22–23.

Towell, Richard, Roger Hawkins, and Nives Bazergui. 1996. The development of fluency in advanced learners of French. *Applied Linguistics* 17: 84–119. [CrossRef]

Tranel, Bernard. 1987. *The Sounds of French*. Cambridge: Cambridge University Press.

Wade, Travis, Allard Jongman, and Joan Sereno. 2007. Effects of acoustic variability in the perceptual learning of non-native accented speech sounds. *Phonetica* 64: 122–44. [CrossRef] [PubMed]

Walker, Douglas. 1984. *The Sounds of Canadian French*. Ottawa: University of Ottawa Press.

Walker, Douglas. 2001. *French Sound Structure*. Calgary: University of Calgary Press.

White, Lydia. 2003. *Universal Grammar and Second Language Acquisition*. Cambridge: Cambridge University Press.

Zhang, Yang, Patricia Kuhl, Toshiaki Imada, Paul Iverson, John Pruitt, Erica Stevens, Masaki Kawakatsu, Yoh'ichi Tohkura, and Iku Nemotoc. 2009. Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *NeuroImage* 46: 226–40. [CrossRef] [PubMed]