PHONETIC TRAINING OF L2 SPANISH RHOTICS[*]

Andrew McCandless
University of Toronto

Cross-linguistically, rhotics exhibit a great deal of variation in several parameters. They can vary in manner of articulation, from approximants (as in the English /ɹ/) to fricatives (as in the French and German /ʁ/), to taps (as in the Spanish, Portuguese, Catalan, and Italian /ɾ/) to trills (as in the Spanish, Portuguese, Catalan, and Italian /r/) (Kostakis 2007, Ackerlind and Jones-Kellogg 2011, Colantoni et al. 2015). Also, in place of articulation, rhotics can be retroflex (as in the English /ɹ/), uvular (as in the French and German /ʁ/), or alveolar (as in the Spanish, Portuguese, Catalan, and Italian tap and trill, /ɾ/ and /r/) (Kostakis 2007, Ackerlind and Jones-Kellogg 2011, Colantoni et al. 2015).

The focus of this pilot study is the following: With phonetic training (perception and production training), do low-proficiency L2 Spanish learners with L1 Canadian English achieve more native-like production of the Spanish tap and trill? In the following sections, I will discuss phonological and phonetic characteristics of Spanish and English rhotics, previous findings on variables which may influence L2 production of Spanish rhotics, previous findings on L2 Spanish rhotic perception and perception in general, as well as the essential elements of phonetic training. I will then describe the research question, hypotheses, methodology, results and conclusions for my study.

1. Spanish rhotics

In most varieties of Spanish, there is a phonological contrast in intervocalic position between a voiced alveolar tap and a voiced alveolar trill, as in the following examples:

(1) a. *_caro_ ‘expensive’
    b. _carro_ ‘car’.

Regional variations of both of these rhotics are attested throughout the Spanish-speaking world (Quilis 1999, Face 2006, Hurtado and Estrada 2010, Hualde 2014). However, the tap and trill are most common across different varieties of Spanish, and are the most commonly taught rhotics in L2 Spanish classrooms in North America (Face 2006).

[*] This paper was presented at the 2020 annual conference of the Canadian Linguistic Association, hosted online by Western University, May 30, 2020. The presentation was based on a course assignment for my PhD program in Hispanic Linguistics at University of Toronto, submitted April 21, 2017. The content of sections 2 to 6 is also based on my doctoral dissertation, “A perception and production training study of L2 Spanish rhotics in learners with L1 Canadian English or Canadian French” (in progress). I would like to thank Laura Colantoni and Jeffrey Steele for their help and comments on this paper.

Actes du congrès annuel de l’Association canadienne de linguistique 2020.
Proceedings of the 2020 annual conference of the Canadian Linguistic Association.

1.1 Phonological features

In intervocalic position, the tap and trill contrast. This is the only position where the contrast between rhotics is conveyed orthographically. The tap is represented as <r> and the trill as <rr> (see example 1 above). In all other phonological environments, the two Spanish rhotics either occur in complementary distribution or in free variation. Both rhotics are represented orthographically as <r>.

In onset position (both word-initially and word-medially), only the trill is realized, as in the following examples:

(2) a. **regla* 'rule, ruler'
b. *enredar* 'to tangle, to complicate'.

After a word-initial or word-medial tautosyllabic consonant (a stop or /f/), in general, only the tap is realized, as in the following examples:

(3) a. **grande* 'large'
b. *cerebro* 'brain'
c. *frío* 'cold'
d. *ofrecer* 'to offer, to provide'.

Word-finally, before a vowel, only the tap occurs, as in the following example:

(4) **ser amable* 'to be kind'.

In coda position (both word-medially and word-finally, before a consonant or a pause), the tap is more common, but the trill can also occur, as in the following examples:

(5) a. **harto* 'full, a lot of, fed up'
b. *estar feliz* 'to be happy'
c. *hablar* 'to speak'.

(Quilis 1999; Colantoni and Steele 2005, 2008, 2011; Schwegler et al. 2010; Hualde 2014).

1.2 Phonetic features

The Spanish tap is realized with one brief apico-alveolar articulation, and is a voiced segment. Mean duration for this segment among native speakers is 20 to 25 ms, with a range of 16 to 36 ms. The tap is articulated similarly to the Canadian English flap (a post-tonic, intervocalic allophone of /t/ or /d/, for example, in the words *later* or *ladder*). However, unlike the flap, it is not realized with any overshoot or retraction before the apico-alveolar contact (Colantoni and Steele 2005, 2008, 2011; Face 2006; Weech 2009; Hurtado and Estrada 2010; Rose 2010; Schwegler et al. 2010; Olsen 2012, 2016; Gick et al. 2013; Daidone and Darcy 2014; Hualde 2014; Scarpace 2014; Colantoni et al. 2015).

On the other hand, the Spanish trill is articulated with two to six (mean of three) brief apico-alveolar articulations, with vowel-like segments between the occlusions. Similar to the tap, it is a voiced segment. Its mean duration is 85 ms (15 ms per closure and 18 ms per vocalic element for a three-closure trill), with variation according to position and following vowel. Mean formant measurements for trills (F1, F2, and F3, respectively) are 500 Hz, 1550 Hz, and 2500 Hz (Quilis 1999, Colantoni 2006, Schwegler et al. 2010, Hualde 2014, Colantoni et al. 2015).

The trill is one of the last segments to be acquired in production by L1 Spanish children. Articulatorily and aerodynamically, it is a very complex segment. Articulatorily, it is more complex than a series of taps. Compared to the tap, its production requires a great deal of more muscle tension and more precision in articulatory gestures. Aerodynamically, production of apico-alveolar trills is an example of the Bernoulli effect, in which oropharyngeal pressure increases and airflow decreases during each closure phase, and the reverse occurs during each release phase (Solé 2002).

2. Canadian English rhotics

2.1 Retroflex alveolar approximant

English varies greatly from Spanish in rhotic inventory, phonology, and phonetics. Canadian English, similar to many other varieties of English, has one rhotic phoneme, a retroflex alveolar approximant, /ɹ/.[1] In general, this segment does not exist in Spanish (Colantoni and Steele 2008, Schwegler et al. 2010, Hualde 2014, Colantoni et al. 2015, O'Grady and Archibald 2016). The English rhotic phoneme is represented orthographically as <r> (all positions) or <rr> (optionally, in intervocalic position). It can occur in word-initial singleton onsets (*road*), word-initial clusters (*trace*), intervocalically (*caring*, *arrive*), word-medial clusters (*afraid*), word-final clusters (*fort*), and word-final singleton codas (*car*) (Bradlow et al. 1997, Colantoni and Steele 2008, Scarpace 2014).

Phonetically, the Canadian English /ɹ/ is articulated as a (generally voiced) apico- or dorso-alveolar approximant. In some speakers, /ɹ/ is produced as a retroflex [ɹ], with the apex raising and retracting toward the prepalatal area, and the tongue being concave. In others, it is produced as a bunched [ɹ], with the apex lowering and contracting, the dorsum raising toward the mediopalatal area, and the tongue being convex. Mean duration can vary from 90 to 148 ms, depending on position (intervocalic ≈ word-initial singleton onsets < word-final singleton codas < word-initial clusters). Mean F1, F2, and F3 for this segment, respectively, are 380 Hz, 1200 to 1310 Hz, and 1500 to 1660 Hz (Henly and Sheldon 1986; Stevens 1998; Quilis 1999; Colantoni and Steele 2008; Olsen 2012, 2016; Hualde 2014; Colantoni et al. 2015).

2.2 Alveolar flap

Similar to American English, Canadian English has an alveolar flap allophone, [ɾ], which

[1] Actually (see below), this rhotic can be realized with retroflexion or bunching. However, the term "retroflex alveolar approximant" is most often used to classify it.

is very similar (but not identical) to the Spanish tap. This segment, represented orthographically as <t>, <tt>, <d> or <dd>, occurs only in post-tonic, intervocalic position, as an allophone of /t/ or /d/ (e.g., *city*, *matter*, *edit*, *ladder*). Articulatorily, the Canadian English [ɾ] is a voiced apico-alveolar flap. Its duration varies from 23 to 36 ms. Unlike the Spanish tap, it is produced with overshoot and retraction before the apico-alveolar contact (Face 2006; Colantoni and Steele 2008; Weech 2009; Hurtado and Estrada 2010; Rose 2010; Schwegler et al. 2010; Olsen 2012, 2016; Gick et al. 2013; Daidone and Darcy 2014; Hualde 2014; Scarpace 2014; Colantoni et al. 2015; O'Grady and Archibald 2016).

3. L2 production of Spanish rhotics

3.1 Linguistic variables

Previous studies on L2 production of Spanish rhotics have provided evidence that both linguistic and learner variables may influence rhotic production a great deal. Regarding linguistic variables, previous studies have provided evidence that target-like production for taps may be easier than for trills, since taps are easier to articulate than trills (Face 2006; Johnson 2008; Olsen 2012, 2016). Also, it may be easier to produce rhotics intervocalically than in other positions, due to facilitated airflow and fewer articulatory constraints intervocalically (Colantoni and Steele 2008; Johnson 2008; Hurtado and Estrada 2010; Olsen 2012, 2016). In addition, it may be difficult for L2 learners to acquire simultaneously the phonetic parameters of duration, voicing, and manner for Spanish rhotics in different phonological environments. The order of acquisition of these parameters may vary depending on phonological environment, and they may be acquired in one phonological environment before others (Colantoni and Steele 2008).

Previous research has provided evidence that at the beginning stages of L2 learning, L1 articulatory routines may influence L2 Spanish rhotic production (Olsen 2012). However, as L2 proficiency increases, the influence of these L1 articulatory routines may disappear (Olsen 2016). Acquisition patterns for trills may be categorical, gradient, or fossilized, depending on various aerodynamic measures (i.e., airflow, number of taps per trill, delay between taps, total tap duration, closure duration, time to vent pressure, open quotient) and phonological environment combined (Johnson 2008).

3.2 Learner variables

Regarding learner variables, previous research has provided some evidence that transfer errors may decrease and target-like rhotic productions may increase as L2 Spanish proficiency increases (Face 2006, Johnson 2008). As L2 Spanish proficiency increases, realizations for target trills may shift from L1 rhotics (transfer errors) to taps (developmental errors) to trills (target rhotics) (Face 2006, Johnson 2008, Weech 2009). Previous studies have provided evidence that formal pronunciation instruction (at home or abroad) may influence target-like rhotic production (Hurtado and Estrada 2010) and that living in a Spanish-speaking country for more than one year may benefit target-like production of L2 Spanish rhotics (Weech 2009).

4. L2 perception of Spanish rhotics

Although this study focuses on production of Spanish rhotics, a great deal of research has provided evidence for a link between perception and production. Many L2 phonology models postulate a perception-production link (e.g., Flege 1995, Major 2001). Several theoretical models of perception and production propose a direct or indirect link between the two modalities (e.g., Liberman and Mattingly 1985, Fowler 1986). Also, many neuroscience studies have provided evidence that listening to speech may activate motor, as well as auditory, speech-related cortical areas (e.g., Wilson et al. 2004), and that silently producing speech may activate auditory, as well as motor, speech-related cortical areas (e.g., Price et al. 2011).

Previous research on L2 perception of Spanish rhotics has provided evidence that L2 Spanish learners may discriminate the intervocalic tap and trill with high accuracy (Rose 2010, Daidone and Darcy 2014). However, these learners may have difficulty lexically encoding this contrast (i.e., listening to a word with the tap or the trill and associating the correct rhotic with the correct word), which is associated with perceptual identification (Daidone and Darcy 2014, Scarpace 2014). Encoding this contrast may be easier intervocalically (tap: *pe<u>r</u>o* 'but' vs. trill: *pe<u>rr</u>o* 'dog') than at word boundaries (tap: *pedi<u>r</u> elotes* 'to ask for corn' vs. trill: *pedí <u>r</u>elojes* 'I asked for clocks') (Scarpace 2014).

5. Phonetic training

Previous research has provided evidence for the effectiveness of phonetic training (perception and/or production training) in improving perception and production of a variety of L2 segments in many L1 groups, including L2 Spanish rhotics in L1 American English speakers (Bradlow et al. 1997, 1999; Herd et al. 2013; Kartushina et al. 2015; Sakai 2016). Objectives of phonetic training are to evaluate L2 learners' degree of pre- to post-test improvement, transfer of improvements from the trained modality to the opposite modality (perception to production, or vice versa), generalization of learning from training to new contexts, new speakers and new stimuli, and long-term retention of improvements with training.[2] Training duration can be short-term (one session) or long-term (multiple sessions), with long-term being much more common than short-term training (Logan and Pruitt 1995, Sakai and Moorman 2017).

Improvements with training are most commonly evaluated through a pre- and post-test. The tasks and stimuli at pre- and post-test are identical to each other and similar to those used in training. Results are most commonly measured based on percentage of correct responses at pre- and post-test, but can also be measured based on improvement and gain scores, response times, rating scales or evoked potentials (Logan and Pruitt 1995; Bradlow et al. 1997, 1999; Herd et al. 2013; Kartushina et al. 2015; Sakai 2016; Sakai and Moorman 2017).

Two common perception training and testing tasks used in previous studies are

[2] Typically, long-term retention of improvements with training is investigated several months after post-test (e.g., Bradlow et al. 1999). Due to the severe time limitations of this study (it was completed for a course assignment), it was not possible to assess long-term retention of pre- to post-test improvements.

discrimination tasks, in which two or three stimuli are presented in each trial and the participant differentiates them auditorily, and identification tasks, in which one stimulus is presented in each trial and the participant identifies the stimulus from a closed set of responses (Logan and Pruitt 1995; Bradlow et al. 1997, 1999; Herd et al. 2013; Kartushina et al. 2015; Sakai 2016; Sakai and Moorman 2017). Some production training and testing tasks used in previous studies are elicited production tasks, in which the participant speaks segments or words out loud, reading tasks (individual words or a passage), and picture description tasks (Bradlow et al. 1997, 1999; Herd et al. 2013; Kartushina et al. 2015; Sakai 2016).

Feedback from training is very important to ensure that learning occurs, since the time period for training is limited. The most common (and perhaps, most important) type of feedback for effective learning from training is immediate, or trial-by-trial, feedback. In addition, cumulative feedback (provided after each block of trials or at the end of each session) may also motivate participants to continue with training (although it may be less essential to learning from training than immediate feedback). Feedback most commonly contains information on whether participants' responses are correct or incorrect, but can also include information on their response times. It is only provided during training, not during the pre- or post-test (Logan and Pruitt 1995).

6. Research question and hypotheses

The research question for this study is the following: With phonetic training (i.e., perception and production training), do low-proficiency L2 Spanish learners with L1 Canadian English achieve more native-like production of the Spanish tap and trill? It was hypothesized that with training: (1) These learners would produce intervocalic taps with more native-like duration and manner, but would improve less in voicing, and improvement with training for taps in consonant clusters would be smaller than for intervocalic taps. (2) They would produce intervocalic trills with more native-like duration, voicing and manner, but would improve less for word-initial and syllable-initial, word-medial trills. (3) They would show less improvement for syllable-final, word-medial rhotics and word-final, pre-pausal rhotics than for intervocalic taps and trills. Stimuli for this study consisted of Spanish words with rhotics in seven different phonological environments: intervocalic taps, post-consonantal taps, word-initial trills, intervocalic trills, syllable-initial, word-medial trills, syllable-final, word-medial rhotics, and syllable-final, word-final rhotics.

7. Methodology

7.1 Participants

One L1 Canadian English speaker with beginner L2 Spanish (P003, age 71, female) was recruited for this study.[3] P003 was from a small town in Ontario, near the Greater

[3] This was the minimum number of participants required for the assignment. More participants are needed to evaluate the hypotheses more thoroughly and to have stronger evidence for this study's conclusions.

Toronto Area. She reported to have no more than beginner proficiency in another L2, and normal hearing and dentition. Efforts were made to recruit a control participant for this experiment. However, this latter participant was not available for testing.

7.2 Pre- and post-test task

The pre- and post-test task was a reading task, which consisted of 120 stimuli, in four blocks of 30, in the carrier phrase *Digo* **____** *otra vez* ('I say **____**again'). This task was completed twice (once before and once after training). It included 70 target stimuli (10 per rhotic and position), and 50 distracters, presented in random order. Of the target stimuli, 35 contained a rhotic in a stressed syllable, and the other 35 contained a rhotic in an unstressed syllable (five per rhotic and position for each stress type). Stimuli were real, two- to three-syllable words, and were mainly nouns, verbs and adjectives. Due to the large number of stimuli, P003 paused for 30 seconds after each block to reduce fatigue. There were 10 familiarization items presented at the beginning of the task.

The pre- and post-test tasks, as well as all of the other tasks for this study, were recorded in mono, using a unidirectional microphone (a Zoom H4n Pro), at a sampling frequency of 44 100 Hz and at a sampling rate of 16 bits. These recordings were saved as .wav files. Then, target utterances were extracted from these files and saved as individual sound files (.wav) using Audacity (Audacity Team 1999/2017), and labelled for analysis in Praat (Boersma and Weenink 1992/2015). TextGrids were created and saved for each individual stimulus, following the procedure in Colantoni et al. (2015, pp. 168-174).

7.3 Training

The training task consisted of two phases. In the first phase (perception training), P003 listened to 50 Spanish words, recorded by an adult male native Mexican Spanish speaker, with orthography presented on a computer screen (two repetitions per word, with a three-second interstimulus interval (ISI) between repetitions). This task consisted of 35 target stimuli (five per rhotic and position) and 15 distracters. Of the target stimuli, 16 contained a rhotic in a stressed syllable, and the other 19 contained a rhotic in an unstressed syllable. Stimuli were real, two- to three-syllable words, presented auditorily, in random order, with a desktop computer and external (Bose) speakers.

In the second phase (production training), P003 listened to (twice, with a three-second ISI between presentations) and then, repeated (once) 100 Spanish words (including the same 50 from the first phase), presented in written form on a computer screen. The task included 70 target stimuli (10 per rhotic and position, including the same 35 from the first phase), and 30 distracters (including the same 15 from the first phase). Of the target stimuli, 33 contained a rhotic in a stressed syllable, and the other 37 contained a rhotic in an unstressed syllable. Stimuli were real, two- to three- syllable words, presented in random order, in the same way as in perception training. Verbal feedback was provided for each stimulus (positive reinforcement when target-like, encouragement to repeat again if not target-like). The same Mexican Spanish speaker recorded all non-repeated target stimuli and distracters, and half of the repeated target stimuli (18 items) and distracters (seven items). An adult male native Peninsular Spanish speaker recorded the other repeated target stimuli (17 items) and distracters (eight items).

Following the procedure of Trofimovich and Gatbonton (2006), between the two training stages, P003 answered 20 simple arithmetic problems to erase stimuli in the perception stage from short-term memory. Incorporating this task between the two training phases also reduced the influence of perception on production (see Sakai 2016).

7.4 Testing protocol

Testing took place at the participant's home, in a quiet room. In consultation with P003 (with whom the researcher was well acquainted), it was decided to not provide remuneration. Testing took place in the following order: (1) Consent form and questionnaire (eight minutes), (2) Pre-test task (seven minutes), (3) Perception training (10 minutes), (4) Distracter task (three minutes), (5) Production training (15 minutes), (6) Post-test task (seven minutes). The total duration was approximately 50 minutes, completed in one session. P003 was encouraged to take short breaks to avoid fatigue.

8. Data analysis

Using Praat (Boersma and Weenink 1992/2015), the following parameters were measured (following Colantoni and Steele 2008): duration, voicing and manner. Duration and voicing were measured quantitatively (duration in milliseconds, and voicing in milliseconds and as a percentage). For these ratio variables, mean, range (maximum and minimum), and standard deviation (rounded to two decimal places) were calculated.[4] Manner was measured qualitatively, based on waveforms and spectrograms, as approximants, taps, trills or other. For this nominal variable, mode, absolute frequency values and relative frequency values (percentages) were calculated. The results for all three variables were compared within participants (pre- versus post-test for P003) to assess the effects of training on production of Spanish rhotics.

9. Results

9.1 Duration

Regarding target intervocalic taps, there was almost no change in mean duration from pre- to post-test (an increase from 89 to 89.9 ms). However, variability increased from pre- to post-test (pre-test: minimum: 82 ms, maximum: 104 ms, SD: 6.27 ms; post-test: minimum: 77 ms, maximum: 112 ms, SD: 11.01 ms). For target post-consonantal taps, there was little change in mean duration from pre-to post-test (90.9 to 96 ms). Variability increased from pre- to post-test (pre-test: minimum: 81 ms, maximum: 111 ms, SD: 10.54 ms; post-test: minimum: 81 ms, maximum: 112 ms, SD: 14.15 ms). (Figure 1)

[4] Because of the extremely small sample size for this pilot study, it was not possible to conduct inferential statistical tests (e.g., t-tests, ANOVAs). More participants are needed in order for this statistical testing to be conducted meaningfully.

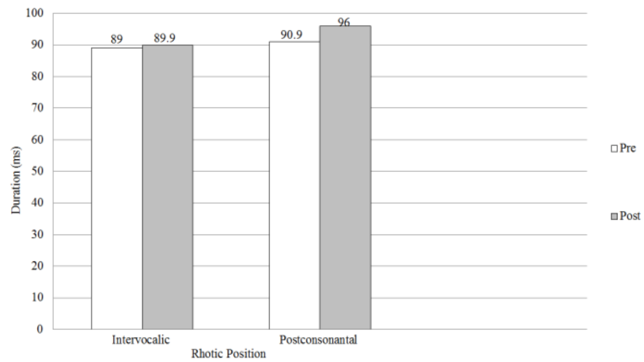

Figure 1. Mean duration of target taps, by position and test.

For target word-initial trills, mean duration increased substantially from pre- to post-test (89.4 to 101.6 ms). Variability increased substantially from pre- to post-test (pre-test: minimum: 83 ms, maximum: 102 ms, SD: 5.97 ms; post-test: minimum: 79 ms, maximum: 116 ms, SD: 11.74 ms). For target intervocalic trills, mean duration increased substantially from pre- to post-test (86.4 to 101.8 ms). Variability increased from pre- to post-test (pre-test: minimum: 79 ms, maximum: 93 ms, SD: 4.17 ms; post-test: minimum: 93 ms, maximum: 113 ms, SD: 8.65 ms). For target syllable-initial, word-medial trills, mean duration increased somewhat from pre- to post-test (87.9 to 97.1 ms). Variability increased from pre- to post-test (pre-test: minimum: 77 ms, maximum: 102 ms, SD: 9.10 ms; post-test: minimum: 82 ms, maximum: 116 ms, SD: 13.57 ms). (Figure 2)

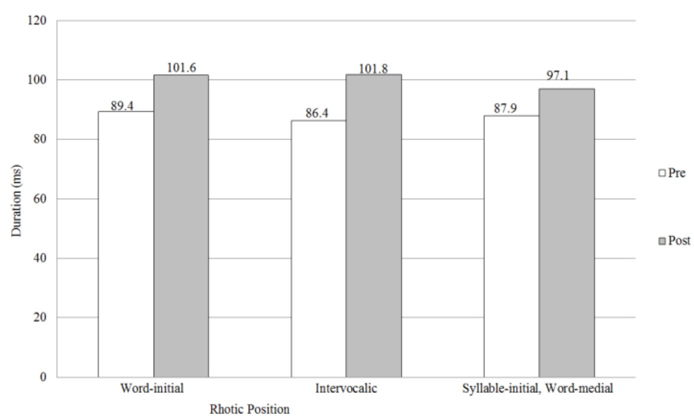

Figure 2. Mean duration of target trills, by position and test.

For syllable-final, word-medial rhotics, mean duration increased somewhat between pre- and post-test (84.3 to 91.9 ms). Variability increased from pre- to post-test (pre-test: minimum: 77 ms, maximum: 90 ms, SD: 4.35 ms; post-test: minimum: 79 ms, maximum: 100 ms, SD: 7.19 ms). For syllable-final, word-final rhotics, mean duration remained almost unchanged from pre- to post-test (91.8 to 91.3 ms). Variability increased from pre- to post-test (pre-test: minimum: 86 ms, maximum: 102 ms, SD: 4.73 ms; post-test: minimum: 83 ms, maximum: 100 ms, SD: 7.82 ms). (Figure 3)

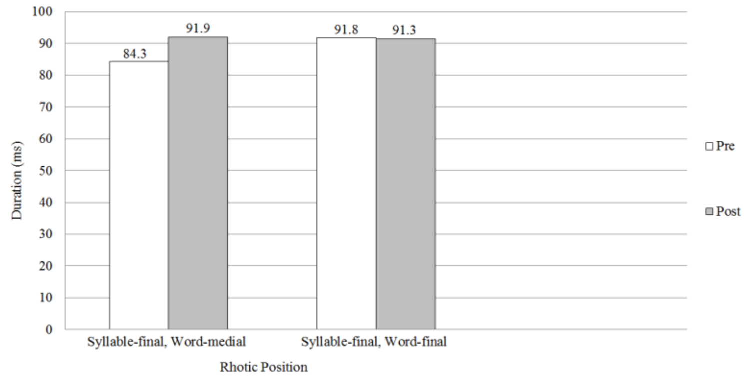

Figure 3. Mean duration of optional taps or trills, by position and test.

9.2 Voicing

For target intervocalic taps, mean voicing increased substantially from pre- to post-test (57.4 to 80.1 ms, 64.4% to 88.9%). Variability (in milliseconds, but not as a percentage) increased substantially from pre- to post-test (pre-test: minimum: 53 ms, 60%, maximum: 67 ms, 73%, SD: 4.10 ms, 4.09%; post-test: minimum: 65 ms, 83%, maximum: 97 ms, 95%, SD: 10.57 ms, 4.15%). For target post-consonantal taps, mean voicing increased substantially from pre- to post-test (59.6 to 84.4 ms, 66% to 87.9%). Variability increased in milliseconds, but decreased as a percentage (pre-test: minimum: 38 ms, 44%, maximum: 77 ms, 78%, SD: 10.06 ms, 10.08%; post-test: minimum: 70 ms, 79%, maximum: 105 ms, 95%, SD: 13.56 ms, 6.31%). (Figure 4)

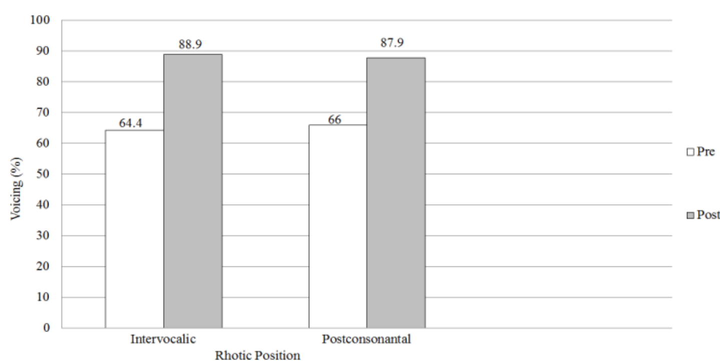

Figure 4. Mean voicing of target taps, by position and test.

For target word-initial trills, mean voicing increased substantially from pre- to post-test (57.3 to 90.7 ms, 64% to 88.9%). Variability increased substantially in milliseconds, but decreased slightly as a percentage (pre-test: minimum: 50 ms, 60%, maximum: 65 ms, 70%, SD: 5.46 ms, 2.98%; post-test: minimum: 67 ms, 84%, maximum: 106 ms, 92%, SD: 11.42 ms, 2.42%). For target intervocalic trills, mean voicing increased substantially from pre- to post-test (55.6 to 91.3 ms, 64.4% to 90%). Variability increased in milliseconds, but decreased as a percentage (pre-test: minimum: 50 ms, 57%, maximum: 63 ms, 68%, SD: 4.53 ms, 4.06%; post-test: minimum: 80 ms, 86%, maximum: 103 ms, 94%, SD: 7.85 ms, 2.59%). For target syllable-initial, word-medial trills, mean voicing increased substantially from pre- to post-test (54.4 to 86.7 ms, 61.9% to 88.9%). Variability increased substantially in milliseconds and slightly increased as a percentage (pre-test: minimum: 47 ms, 58%, maximum: 65 ms, 65%, SD: 5.34 ms, 2.96%; post-test: minimum: 68 ms, 81%, maximum: 106 ms, 92%, SD: 13.27 ms, 3.53%). (Figure 5)

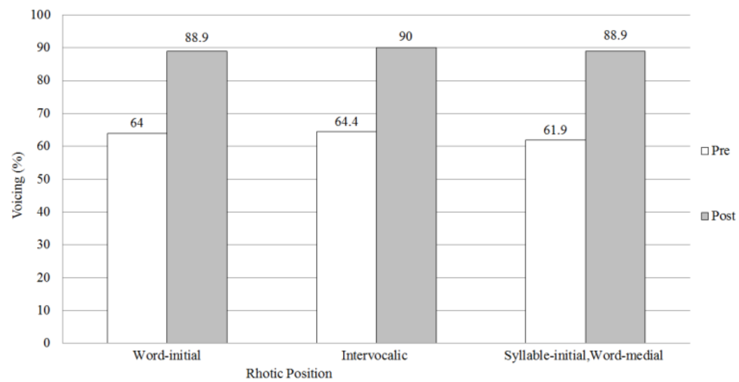

Figure 5. Mean voicing of target trills, by position and test.

Regarding syllable-final, word-medial rhotics, mean voicing increased substantially from pre- to post-test (49.5 to 80.1 ms, 63.4% to 87.3%). Variability showed almost no change in milliseconds. However, it slightly increased as a percentage (pre-test: minimum: 48 ms, 57%, maximum: 60 ms, 68%, SD: 5.67 ms, 2.82%; post-test: minimum: 71 ms, 83%, maximum: 88 ms, 95%, SD: 5.69 ms, 4.03%). In addition, regarding syllable-final, word-final rhotics, mean voicing increased substantially from pre- to post-test (55.3 to 78 ms, 60.2% to 84.5%). Variability increased in milliseconds and slightly increased as a percentage (pre-test: minimum: 48 ms, 56%, maximum: 61 ms, 64%, SD: 4.36 ms, 3.49%; post-test: minimum: 66 ms, 77%, maximum: 87 ms, 89%, SD: 7.96 ms, 3.92%). (Figure 6)

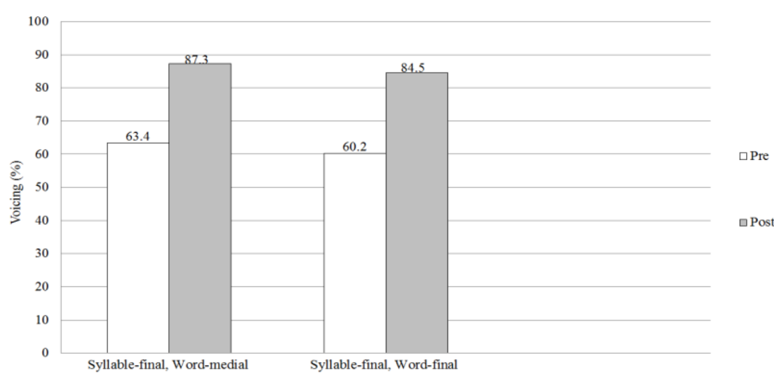

Figure 6. Mean voicing of optional taps or trills, by position and test.

9.3 Manner

For target intervocalic taps, manner changed from approximants at pre-test (10/10, 100%) to a mix of approximants and trills at post-test (5/10, 50%, for both realizations). For target post-consonantal taps, it changed from approximants at pre-test (10/10, 100%) to predominantly trills at post-test (9/10, 90%), with one approximant (10%). (Figure 7)

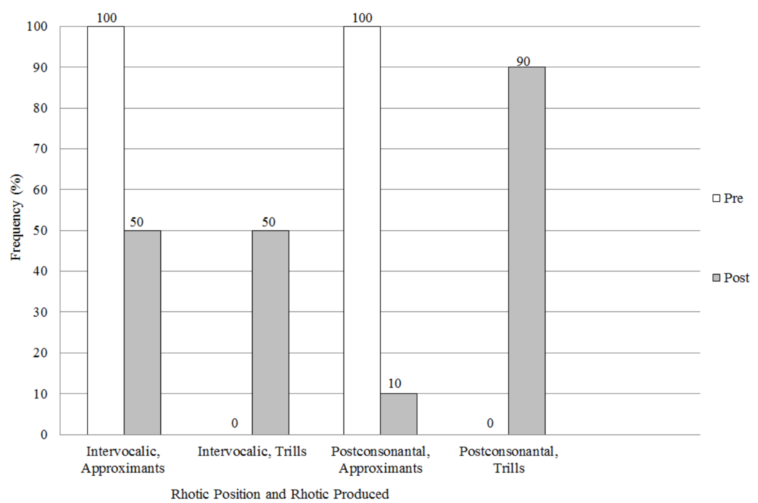

Figure 7. Manner for target taps, by position, rhotic produced, and test.

For target word-initial, intervocalic, and syllable-initial, word-medial trills, manner changed from approximants at pre-test (10/10, 100%) to trills at post-test (10/10, 100%). (Figure 8)

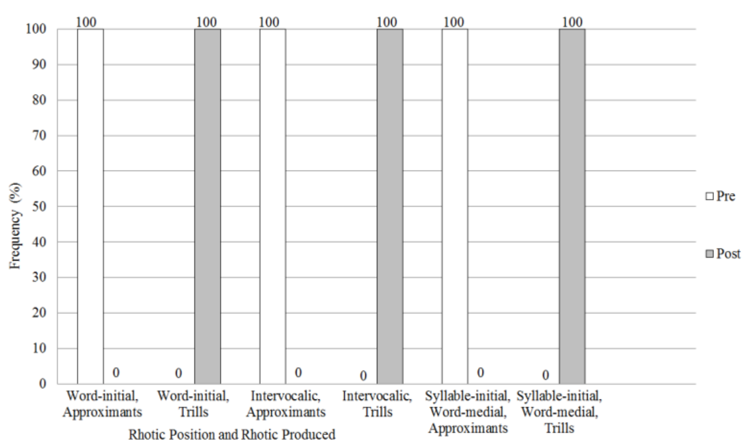

Figure 8. Manner for target trills, by position, rhotic produced, and test.

For syllable-final, word-medial, and syllable-final, word-final rhotics, manner changed from approximants at pre-test (10/10, 100%), to a mix of approximants and trills at post-test (6/10 and 4/10, 60% and 40%, respectively). (Figure 9)

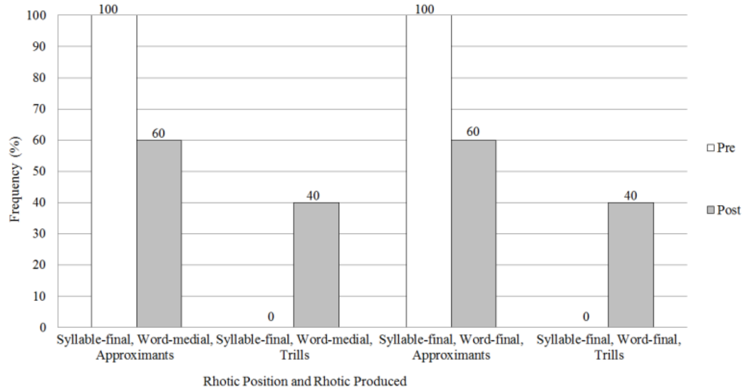

Figure 9. Manner for optional taps or trills, by position, rhotic produced, and test.

10. Discussion

10.1 Summary of results

For target taps (intervocalic and post-consonantal), there were no pre- to post-test improvements observed in duration as a result of training. However, improvements were observed in both voicing and manner, in both phonological environments (with trills co-occurring with approximants intervocalically, and almost replacing them post-consonantally).[5] For target trills (word-initial, intervocalic, and syllable-initial, word-medial), as well as for optional taps or trills (syllable-final, word-medial, and syllable-final, word-final), duration was native-like at pre-test and remained such at post-test. Pre- to post-test improvements were observed in voicing and manner, for target trills and optional taps or trills, in all phonological environments examined. (Table 1)

Table 1. Pre- to post-test improvements (=: target at pre-test, ✓: improvement toward target productions, ≈: improvement toward near-target productions, X: no improvement)

	Rhotic Type		
Parameter	Target Tap	Target Trill	Optional Tap/Trill
Duration	X	=	=
Voicing	✓	✓	✓
Manner	≈	✓	✓

10.2 Trills vs. taps

P003 showed relatively successful pre- to post-test improvement on all parameters for trills (and optional taps or trills), in all phonological environments investigated, perhaps since she may have perceived the trill as a different segment from the English alveolar

[5] Native Spanish speakers sometimes realize target taps as trills, and vice versa (Herd et al. 2013).

approximant or the Canadian English flap. As previously stated, the Spanish apico-alveolar trill is not similar to any segment in Canadian English. Following Flege's Speech Learning Model (1995), the trill is easy to differentiate from the L1 Canadian English approximant and flap categories, and, therefore, easy to produce in a native-like way.

Relatively less pre- to post-test improvement was observed for duration and manner of taps, in all phonological environments investigated. Since the tap is perceptually more similar to the English flap than the trill is (Hurtado and Estrada 2010), taps may have been perceived as equivalent to trills (which are longer in duration than and have a different manner of articulation from the tap), and therefore, may have been produced with longer than target-like duration, and as trills.

10.3 Future work

Several follow-up steps are needed for this study. First, it should be conducted on more L1 Canadian English participants, to evaluate the hypotheses more thoroughly and have stronger evidence for its conclusions. In addition, perception testing should be conducted, since L2 Spanish learners may have difficulty acquiring rhotics in perception (Daidone and Darcy 2014), as well as production (Face 2006). Furthermore, the study should be conducted on participants with other L1s (e.g., French, Portuguese, Italian, German), to investigate differences in rhotic improvement with training as a function of L1.

11. Conclusion

This study has provided evidence that with a short amount of phonetic training, beginner L2 Spanish learners with L1 Canadian English may show improvement in their productions of Spanish rhotics. These learners may show more improvement for trills and optional taps or trills than for taps, since trills are perceptually distinct from any Canadian English segment, while taps are similar to the Canadian English flap. Order of acquisition of parameters may vary by rhotic type (see Colantoni and Steele 2008). In these L2 learners, for taps, voicing may improve before manner and duration, while for trills and optional taps or trills, duration may improve before manner and voicing.

References

Ackerlind, Sheila R., and Rebecca Jones-Kellogg. 2011. *Portuguese: A reference manual*. Austin, TX: University of Texas Press.

Audacity Team. 1999. Audacity®: Free Audio Editor and Recorder [Computer application]. Version 2.1.3, retrieved 17 March 2017 from https://audacityteam.org/.

Boersma, Paul, and David Weenink. 1992. Praat: doing phonetics by computer [Computer program]. Version 5.4.04, retrieved 12 January 2015 from http://www.praat.org/.

Bradlow, Ann R., David B. Pisoni, Reiko Akahane-Yamada, and Yoh'ichi Tohkura. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America* 101(4): 2299–2310.

Bradlow, Ann R., Reiko Akahane-Yamada, David B. Pisoni, and Yoh'ichi Tohkura. 1999. Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics* 61(5): 977–985.

Colantoni, Laura. 2006. Increasing periodicity to reduce similarity: An acoustic account of deassibilation in rhotics. In *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*, ed. Manuel Díaz-Campos, 22–34. Somerville, MA: Cascadilla Proceedings Project.

Colantoni, Laura, and Jeffrey Steele. 2005. Liquid asymmetries in French and Spanish. In *Toronto Working Papers in Linguistics* 24: *Special Issue on Similarity in Phonology*, ed. Chiara Frigeni, Manami Hirayama, and Sara Mackenzie, 1–14.

Colantoni, Laura, and Jeffrey Steele. 2008. Integrating articulatory constraints into models of second language phonological acquisition. *Applied Psycholinguistics* 29(3): 489–534.

Colantoni, Laura, and Jeffrey Steele. 2011. Synchronic evidence of a diachronic change: Voicing and duration in French and Spanish stop-liquid clusters. *The Canadian Journal of Linguistics/La revue canadienne de linguistique* 56(2): 147–177.

Colantoni, Laura, Jeffrey Steele, and Paola Escudero. 2015. *Second language speech: Theory and practice*. Cambridge, UK: Cambridge University Press.

Daidone, Danielle, and Isabelle Darcy. 2014. Quierro comprar una guitara: Lexical encoding of the tap and trill by L2 learners of Spanish. In *Selected Proceedings of the 2012 Second Language Research Forum*, ed. Ryan T. Miller, Katherine I. Martin, Chelsea M. Eddington, Ashlie Henery, Nausica Marcos Miguel, Alison M. Tseng, Alba Tuninetti, and Daniel Walter, 39–50. Somerville, MA: Cascadilla Proceedings Project.

Face, Timothy L. 2006. Intervocalic rhotic pronunciation by adult learners of Spanish as a second language. In *Selected Proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, ed. Carol A. Klee and Timothy L. Face, 47–58. Somerville, MA: Cascadilla Proceedings Project.

Flege, James E. 1995. Second language speech learning: Theory, findings and problems. In *Speech perception and linguistic experience: Issues in cross-language research*, ed. Winifred Strange, 233–277. Timonium, MD: York Press.

Fowler, Carol A. 1986. An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* 14(1): 3–28.

Gick, Bryan, Ian Wilson, and Donald Derrick. 2013. *Articulatory phonetics.* West Sussex, UK: Wiley-Blackwell.

Henly, Elizabeth, and Amy Sheldon. 1986. Duration and context effects on the perception of English /r/ and /l/: A comparison of Cantonese and Japanese speakers. *Language Learning* 36(4), 505-521.

Herd, Wendy, Allard Jongman, and Joan Sereno. 2013. Perceptual and production training of intervocalic /d, ɾ, r/ in American English learners of Spanish. *The Journal of the Acoustical Society of America* 133(6): 4247–4254.

Hualde, José I. 2014. *Los sonidos del español*. Cambridge, UK: Cambridge University Press.

Hurtado, Luz M., and Chelsea Estrada. 2010. Factors influencing the second language acquisition of Spanish vibrants. *The Modern Language Journal* 94(1): 74–86.

Johnson, Keith E. 2008. Second language acquisition of the Spanish multiple vibrant consonant. Doctoral dissertation, The University of Arizona.

Kartushina, Natalia, Alexis Hervais-Adelman, Ulrich H. Frauenfelder, and Narly Golestani. 2015. The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America* 138(2): 817–832.

Kostakis, Andrew. 2007. More on the origin of uvular [ʀ]: Phonetic and sociolinguistic motivations. In *Indiana University Linguistics Club Working Papers* 7(1), ed. Catalina Mendez-Vallejo and Ken De Jong, Article 2.

Liberman, Alvin M., and Ignatius G. Mattingly. 1985. The motor theory of speech perception revised. *Cognition* 21(1): 1–36.

Logan, John S., and John S. Pruitt. 1995. Methodological issues in training listeners to perceive non-native phonemes. In *Speech perception and linguistic experience: Issues in cross-language research*, ed. Winifred Strange, 233–277. Timonium, MD: York Press.

Major, Roy C. 2001. *Foreign accent: The ontogeny and phylogeny of second language phonology.* Mahwah, NJ: Lawrence Erlbaum.

O'Grady, William, and John Archibald. 2016. *Contemporary linguistic analysis: An introduction*. Don Mills, ON: Pearson Canada.

Olsen, Michael K. 2012. The L2 acquisition of Spanish rhotics by L1 English speakers: The effect of L1 articulatory routines and phonetic context for allophonic variation. *Hispania* 95(1): 65–82.

Olsen, Michael K. 2016. Limitations of the influence of English phonetics and phonology on L2 Spanish rhotics. *Borealis: An International Journal of Hispanic Linguistics* 5(2): 313–331.

Price, Cathy J., Jenny T. Crinion, and Mairéad MacSweeney. 2011. A generative model of speech production in Broca's and Wernicke's areas. *Frontiers in Psychology* 2: Article 237.

Quilis, Antonio. 1999. *Tratado de fonología y fonética españolas.* 2nd ed. Madrid, Spain: Gredos.

Rose, Marda. 2010. Differences in discriminating L2 consonants: A comparison of Spanish taps and trills. In *Selected Proceedings of the 2008 Second Language Research Forum*, ed. Matthew T. Prior, Yukiko Watanabe, and Sang-Ki Lee, 181–196. Somerville, MA: Cascadilla Proceedings Project.

Sakai, Mari. 2016. (Dis)Connecting perception and production: Training adult native speakers of Spanish on the English /i/-/ɪ/ distinction. Doctoral dissertation, Georgetown University.

Sakai, Mari, and Colleen Moorman. 2017. Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics* 39(1): 1–38.

Scarpace, Daniel. 2014. The acquisition of the tap/trill contrast within and across words in Spanish. In *Concordia Working Papers in Applied Linguistics* 5: *Proceedings of the International Symposium on the Acquisition of Second Language Speech*, ed. Walcir Cardoso and Pavel Trofimovich, 580–596.

Schwegler, Armin, Juergen Kempff, and Ana Ameal-Guerra. 2010. *Fonética y fonología españolas*. 4th ed. Hoboken, NJ: John Wiley & Sons.

Solé, Maria-Josep. 2002. Aerodynamic characteristics of trills and phonological patterning. *Journal of Phonetics* 30(4): 655–688.

Stevens, Kenneth N. 1998. *Acoustic phonetics*. Cambridge, MA: MIT Press.

Trofimovich, Pavel, and Elizabeth Gatbonton. 2006. Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *The Modern Language Journal* 90(iv): 519–535.

Weech, Andrew M. 2009. Second language acquisition of the Spanish tap and trill in a contact learning environment. Master's thesis, Brigham Young University.

Wilson, Stephen M., Ayşe P. Saygin, Martin I. Sereno, and Marco Iacoboni. 2004. Listening to speech activates motor areas involved in speech perception. *Nature Neuroscience* 7(7): 701–702.