RHOTIC VARIATION AND CONTRAST IN VERACRUZ MEXICAN SPANISH

VARIACIÓN Y CONTRASTE EN LAS RÓTICAS DEL ESPAÑOL MEXICANO DE VERACRUZ

TRAVIS G. BRADLEY
University of California, Davis
tgbradley@ucdavis.edu

ERIK W. WILLIS
Indiana University, Bloomington
ewwillis@indiana.edu

Artículo recibido el día: 14/09/2011
Artículo aceptado definitivamente el día: 09/05/2012
Estudios de Fonética Experimental, ISSN 1575-5533, XXI, 2012, pp. 43-74

ABSTRACT

Phonetic studies of Spanish rhotics report a wide range of allophonic variants of the syllable-initial trill /r/, which raises the question of whether the intervocalic contrast between /r/ and the tap /ɾ/ has been neutralized in many dialects. This study presents a spectrographic analysis of syllable-initial rhotics as produced by ten speakers of Veracruz Mexican Spanish in a guided, semi-spontaneous speech task. Trills that show a reduction in the degree of lingual trilling usually contain an approximant phase following one or two lingual contacts, which we represent as [ɾɹ] or [rɹ] in narrow transcription. Intervocalic taps show both reduction and elision, but those with a measurable contact are short enough to maintain an acoustic difference with the longer allophones of /r/. Taken with recent studies of rhotics in Dominican Spanish, these findings suggest that the contrast between /r/ and /ɾ/ can be maintained in terms of overall segmental duration even when there is no difference in the number of lingual contacts.

Keywords: *rhotics, spectrographic analysis, Veracruz Mexican Spanish.*

RESUMEN

En los estudios fonéticos de las consonantes róticas del español, se han comprobado muchas variantes alofónicas de la vibrante /r/ múltiple en posición inicial de sílaba, lo cual plantea la pregunta de si el contraste intervocálico entre la /r/ múltiple y la /ɾ/ simple se ha neutralizado en muchos dialectos. Este estudio presenta un análisis espectrográfico de las vibrantes iniciales de sílaba producidas por diez hablantes del español de Veracruz, México en una tarea guiada de habla semi-espontánea. Las vibrantes múltiples que muestran una reducción del grado de vibración de la lengua típicamente tienen una fase aproximante después de uno o dos contactos linguales, lo cual representamos como [ɾɹ] o [rɹ] en transcripción estrecha. Las vibrantes simples intervocálicas muestran reducción y elisión, pero las que tienen un contacto medible son lo suficientemente cortas como para mantener una diferencia acústica con los alófonos más largos de /r/. Tomados en conjunto con estudios recientes de las róticas del español dominicano, estos resultados sugieren que el contraste entre /r/ y /ɾ/ se puede mantener a base de una diferencia de duración segmental, aun cuando no hay diferencia en el número de contactos linguales.

Palabras clave: *róticas, análisis espectrográfico, español de Veracruz, México.*

1. INTRODUCTION

Understanding the behavior and cross-linguistic patterning of rhotics, or *r*-type sounds, is an important goal of phonological research. Rhotics form such a heterogeneous class of sounds that there is little consensus in the literature as to what their defining phonetic features are. Within this class there is wide variation in manner and place of articulation, which is observable across languages, dialects, and speech styles. Despite their phonetic variation, rhotics seem to form a single phonological class in both synchronic and diachronic alternations. The strongest evidence for phonological classhood comes from the observation that *rhotics of one type often alternate with other rhotics* (Ladefoged and Maddieson, 1996: 216).

Spanish provides an ideal case study of rhotic variation, which has been documented extensively in contemporary dialects. Normative Spanish has two contrastive rhotic segments: the *vibrante simple,* or tap [ɾ], and the *vibrante múltiple,* or trill [r]. Both are usually voiced and alveolar, but they differ in articulation and number of lingual contacts (Blecua, 2001; Hualde, 2005: 181; Martínez Celdrán and Fernández Planas, 2007: 143-161; Recasens, 1991; Recasens and Pallarès, 1999; Solé, 2002). The tap involves a ballistic articulatory gesture, which produces a single brief and often incomplete contact between the tongue tip and the alveolar ridge. The trill requires a more precise gesture to create the aerodynamic conditions for passive vibration of the tongue tip, which produces between two to five brief contacts. Acoustic measurements indicate differences in overall segment duration. Blecua (2001) reports average durations of 23 ms and 64 ms for the Peninsular Spanish tap and the trill, respectively, which are comparable to averages reported by Quilis (1993), 20 ms and 85 ms.

The examples in (1) illustrate the general distribution of rhotics in normative Spanish, based on Hualde's (2005: 183) description. The phonological contrast is found only in word-medial intervocalic position (1a) and is supported by fewer than 30 minimal pairs. The rhotics are in complementary distribution in other prevocalic contexts. Trills appear in syllable-initial position, both word-initially and after a heterosyllabic consonant (1b). Taps appear in the second position of a complex onset and word-finally when resyllabified into the onset of a following vowel-initial word (1c). The rhotics are in non-contrastive variation in syllable-final position (1d).

(1)	a.	Contrast tap /ɾ/ vs. trill /r/	V._V /kaɾo/ vs. /karo/	Intervocalic
	b.	Only trill /r/	#_ /roka/	Word-initial
			C._ /alrededor/, /enredo/, /israelita/	After a heterosyllabic consonant
	c.	Only tap /ɾ/	C_ /bɾoma/, /gɾamo/	After a tautosyllabic consonant
			V._#V /seɾ amigos/	Word-final before a vowel
	d.	Variable rhotic (most commonly [ɾ])	V_C /paɾte/ [paɾte] ~ [parte]	Before a consonant
			V_#C /seɾ poeta/	Word-final before a consonant
			V__## /seɾ o no seɾ/	Word-final before a pause

Beyond normative Spanish, the phonetic realization of syllable-initial trills in contexts (1a,b) serves as a delimiting feature among dialects. Experimental studies have documented many phonetic variants that differ in manner of articulation (approximants, fricatives, and taps), place of articulation (coronal, velar, and uvular), and laryngeal setting (voiced, voiceless, breathy voiced) (Blecua, 2001; Bradley, 2006; Colantoni, 2001, 2006a,b; Díaz-Campos, 2008; Hammond, 1999, 2000a,b, 2006; Henriksen and Willis, 2010; Kouznetsov and Pamies, 2008, 2011; Lewis, 2004; Sessarego, 2011; Simonet and Carrasco, 2006; Willis, 2006, 2007; Willis and Bradley, 2008).

Such a wide range of allophonic variation raises the question of whether the contrast in (1a) is subject to neutralization, at least for some dialects or speakers. Based on recordings of a text read aloud by 229 speakers of different Latin American and Peninsular Spanish dialects, Hammond (1999: 136) argues that the three-contact trill prescribed by the Real Academia Española does not occur in normal discourse in the speech of the vast majority of native speakers, and he even goes so far as to claim that *a neutralization of* [ɾ] *and* [r] *has occurred in many dialects in intervocalic environments* (1999: 147). The strength of this claim is somewhat diminished by the fact that in his study only tokens with at least three lingual contacts were counted as normative productions (see Hammond 1999: 142-143). However, neutralization remains a real possibility and has in fact been documented for Judeo-Spanish varieties spoken in the Balkan Peninsula and Rumania. In these dialects, the trill has been lost in favor of the tap, which is the only remaining rhotic in the phonological inventory, e.g. [ˈka.ɾu] < [ˈka.ru], [ˈtje.ɾa] < [ˈtje.ra], [ˈɾo.sa] < [ˈro.sa] (Quintana, 2006: 84-85, 376).

Given the range of phonetic variation in rhotics attested across Spanish dialects and across languages in general, a more likely scenario is that even when trills are produced with just one lingual contact, additional phonetic traces of the original segment can persist in speech production and potentially serve as cues to the underlying contrast. Empirical support for this view comes from acoustic studies by Kouznetsov and Pamies (2008, 2011) comparing Russian and Peninsular Spanish trills. They report that trills in both languages can be realized as a single contact followed by a vocalic component before the following vowel. However, Spanish phonemic taps lack the vocalic component before a following vowel and also show a higher F2 than trills, suggesting a more anterior constriction location for /ɾ/.

The present study reports novel acoustic data on rhotic variation in the Spanish spoken in Veracruz, Mexico. The goals of our empirical investigation are to document the phonetic realizations of syllable-initial rhotics and to identify any acoustic differences between taps and trills in the contrastive intervocalic position. A comparison with recent studies on rhotic variation in Dominican Spanish then leads us to consider the articulatory and aerodynamic factors that are responsible for the patterning of syllable-initial rhotics in both Spanish varieties.

2. METHODOLOGY

Previous studies of the tap and trill in Spanish have posited various allophones based on impressionistic articulatory distinctions or auditory inspection (Hammond, 1999, 2000a,b, 2006; Zamora and Guitart, 1988; see also Hualde, 2005: 181-188; Lipski, 1994, and the studies cited therein). In the present study, we limit ourselves to an acoustic description of the signal based on both waveform and spectrographic measurements.

2.1. Speakers

Ten university and professional degree students (five males, five females) from Veracruz, Mexico, served as informants and were recorded in May, 2005. They were native speakers of the dialect and were recorded within the dialectal region. Speakers gave their informed consent prior to participating in the study[1].

2.2. Data elicitation

Each speaker was first engaged in an informal interview about him or herself to elicit natural conversation before performing several linguistic tasks. The data examined in the present study were drawn from a task in which speakers produced a narrative telling of the wordless children's picture book by Mayer (1969), *Frog Where Are You?* This story is about a little boy, a *perro,* and a *rana,* which induced multiple productions of word-initial and medial trills, as well as word-medial taps in words such as *pero* and *parece.* Informants were recorded directly onto a laptop computer using CoolEdit, a USBpre external sound card, and a head-mounted Shure 512 microphone. The speech was sampled at 44.1K, and syllable-initial rhotic tokens were extracted and analyzed with PRAAT, an acoustic software analysis program (Boersma and Weenink, 2001). The guided, semi-spontaneous narrative task yielded a total of 555 tokens, consisting of 216 phonemic taps (39%, 21.6 average per speaker) and 339 phonemic trills (61%, 33.9 average per speaker).

2.3. Acoustic measurements

There were several cues associated with the VMS tap, including a reduction in the waveform amplitude, as well as variation and fluctuation in the formant structure. In many cases, a clear reduction in intensity was visible in the waveform at the location of the tap, which made it possible to measure the acoustic duration of the lingual contact. However, in quieter speech, tap tokens in isolated words or in

[1] Informed consent was sought and obtained in accordance with New Mexico State University Institutional Review Board (IRB) policy (Application #5866).

words at the end of an utterance in some cases did not demonstrate any appreciable variation in amplitude. In other cases, the only acoustic evidence for the tap came from variation in the formant tracker. Depending on the formant settings of the editor window within PRAAT, there was often a breakdown in the formant movements associated with the tap, producing a crossover of F3 and F4 as well as a spurious emergence of additional formant tracks between the third and fourth formants[2].

The VMS trills posed some challenges for acoustic analysis as compared to normative Spanish trills due to a large number of productions with considerable *r*-coloring of the vowel following the contact(s) of the trill segment. We considered this vocalic *r*-coloring to be part of the trill[3]. However, in some cases we could hear the *r*-coloring but not objectively identify the acoustic cues to determine a precise boundary for the segment. Therefore, we chose to err on the side of caution and incorporated into the duration of the trill segment only that portion of the token that had identifiable cues. The cues which we used to determine the segmental boundaries of the trill are as follows: a reduction in the amplitude of the waveform, the presence of lingual contacts, a sharp transition in formant structure (primarily of F3 and F4), and changes in intensity. Not all cues were present in all productions, but typically a combination of two or more cues was present. We also confirmed the *r*-coloring in the spectrogram with auditory inspection.

Given the spontaneous, open-ended nature of the narrative task, it was not feasible to manipulate independent linguistic variables beforehand (e.g. vowel context, stress, position within word, etc.) or to control the number of tokens that would be elicited. While this task had the advantage of more closely approximating natural speech, the disadvantage was that many speakers produced an unbalanced number of tokens across different conditions. Furthermore, inter-speaker differences in

[2] For the female speakers, the maximum formant value was set at 5500 Hz and the number of formants at five, with all other values set to standard. The formant tracker did not yield as many tracking cues when the number of formants was set to four; however, there was often a fluctuation in F4 associated with the tap, and this fluctuation was more evident with high and low flanking vowels.

[3] The decision to include the *r*-colored portion of the segment as part of the duration of the trill is debatable. Past research has indicated that Spanish trills can exhibit considerable variation, as in the case of assibilation, pre-breathy voicing, and uvularization, to name a few. Therefore, we consider it appropriate to recognize this additional variation as part of the trill proper, as the *r*-coloring derives directly from this segment.

standard deviation suggested that non-normal distributions of duration measurements around the mean were quite common. A larger pool of speakers and sample of tokens would likely resolve these shortcomings. Given these constraints, we employed nonparametric tests (chi-square and Mann-Whitney's *U*) in an attempt to determine some preliminary trends in the available data. Future investigations of VMS rhotics should include more speakers, a wider range of speech tasks and linguistic variables, and a more controlled experimental design in order to ascertain the degree to which our results are generalizable to the larger VMS speech community.

3. RESULTS

3.1. Taps

The phonological tap was produced with a number of phonetic variants in the VMS corpus. The two most frequently occurring productions, constituting slightly more than half of the total, were an approximant [ɾ̞] with continuous formant structure through the constriction period (figure 1) and a noncontinuant tap [ɾ] showing minimal formant structure and sometimes a release burst (figure 2). Some energy can be observed during the constriction period of the token in figure 2, suggesting that the articulatory closure can be incomplete in noncontinuant taps. For electropalatographic evidence of rhotics with incomplete closure in Argentine and Cuban Spanish, see Kochetov and Colantoni (2011); for Catalan rhotics, see Recasens and Espinosa (2007) and Recasens and Pallarés (1999). The presence of a release burst at the end of some tokens further supports their classification as noncontinuants (albeit with an incomplete seal), as opposed to approximants, which show a greater degree of continuous formant structure through the constriction period but never a release burst. There were several tap variations that were difficult to measure objectively because they lacked reliable acoustic landmarks, showing only a gradually sloping amplitude transition from the surrounding vowels. Nonetheless, these barely visible variants often seemed perceptible to the ear. Tokens in this category, which we label as *perceptual taps* and transcribe as [$^{(\text{ɾ})}$], typically had a slight reduction in the amplitude of the waveform or intensity of F3 or F4, as illustrated in figure 3. Finally, there were also cases of complete elision of the tap segment, exemplified in figure 4.

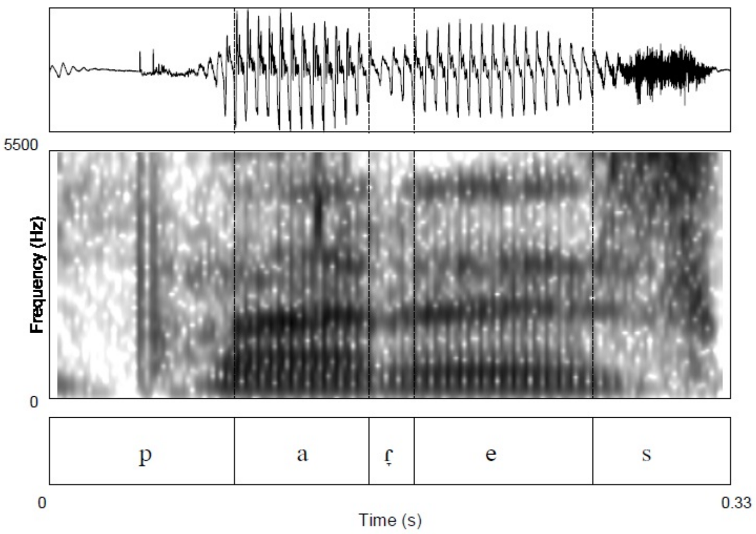

Figure 1. *Approximant tap* [ɾ̞] *in VMS:* parec(e).

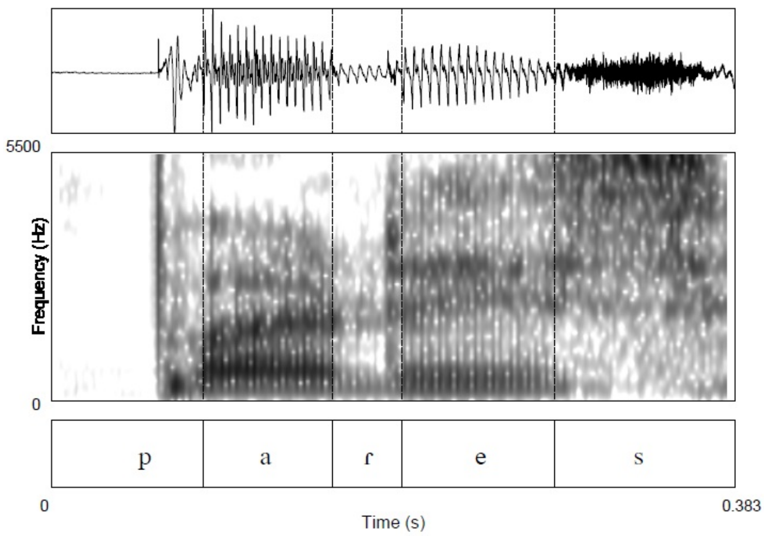

Figure 2. *Noncontinuant tap* [ɾ] *in VMS, with a release burst:* parec(e).

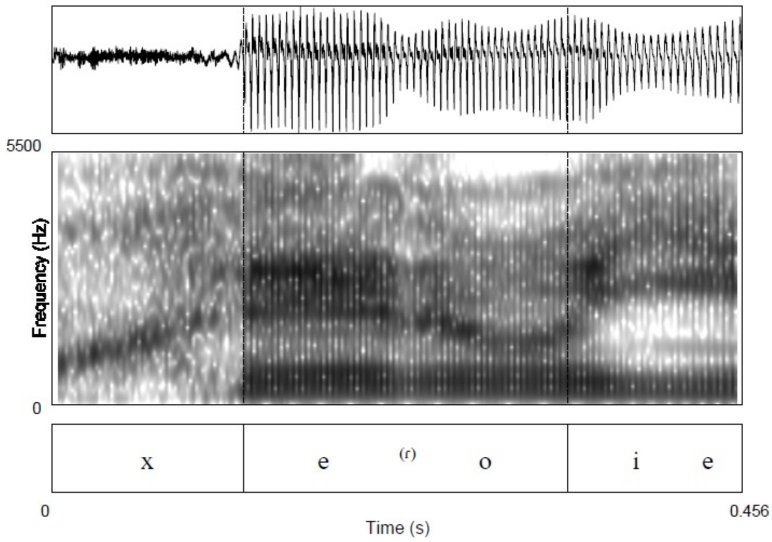

Figure 3. *Perceptual tap [[(ɾ)]] in VMS: (agu)jero y e(l).*

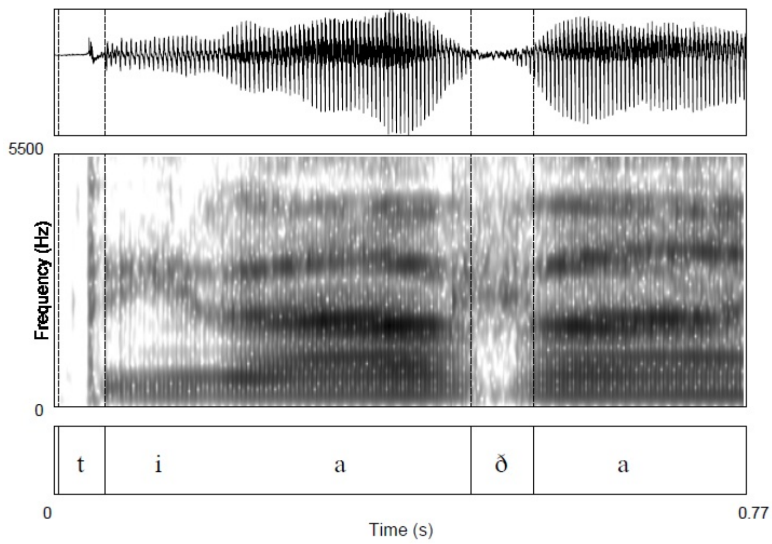

Figure 4. *Complete elision of the tap in VMS:* ti(r)ada.

Because there was so much variation in the realization of the VMS phonological tap, we have collapsed the four principal variants described above into two main groupings, shown in table 1. The two groupings are those tokens with a measurable duration (approximants and noncontinuants) and those that were not measurable due to a lack of consistent and objective cues (perceptual and elided realizations)[4]. Both the perceptual/elided taps and the measureable taps were produced by all speakers, with variation present in the frequency of usage of the two types. The number of tap tokens produced by each speaker varied considerably, ranging from two to 38 tokens. For example, two of the male speakers (Vm4 and Vm5) produced a combined total of only five taps in their story narration, while others produced more than 30 based on the same story. In total, almost half of all taps (48%) were elided or realized as perceptual taps. Speakers showed individual variability in the frequency of these variants, ranging from 28% (speaker Vf1) all the way to 76% (speaker Vf2), with the majority in 30-50% range. A chi-square test of independence was performed to examine the relation between sex and realization of phonological taps as perceptual/elided versus with a measurable contact. The relation between these variables was not significant, χ^2 (1, N = 216) = 1.727, p = 0.189.

As for taps produced with a measurable duration, the VMS speakers typically realized tokens in this category with longer duration than those reported previously for normative taps, and at the same time, these speakers also often produced elided or perceptual taps. Measurable taps, which included approximant and noncontinuant allophones, constituted slightly more than half of all productions at 52%. The average duration of these measurable variants across all speakers was 27.7 ms, which is almost a third longer than the averages reported for normative Spanish taps, 20 ms by Quilis (1993) and 23 ms by Blecua (2001). There was considerable variation in the number of measurable tokens produced and in the duration means across speakers, ranging from 17 ms for speaker Vm1 to 36 ms for Vm3. Only one speaker, Vf4, had an average duration of 21 ms similar to that of normative Spanish taps, but she also produced more than half of her total tap tokens, 14 out of 27, as an elided or perceptual tap. In fact, the VMS speakers as a group participated in segmental reduction, as 48% of all the tokens were produced as elided or perceptual taps. In comparison with normative Spanish characterizations, the VMS taps presented variation suggestive of not only reduction but also segmental

[4] This decision reflects a desire to maintain objectivity and to provide only measurements that are reliably reproducible. We acknowledge that a perceptual tap is in fact an extreme case of approximant realization; however, we have chosen these descriptive labels for ease of discussion.

fortition, as supported by the presence of a release burst on some noncontinuant productions. The speaker with the longest tap duration average, Vm3 at 36 ms, also produced more than a third of tap tokens as a reduced allophone. Due to a lack of tokens illustrating different stress positions, we were unable to make any descriptive observations about the role of stress in the production of taps in VMS.

SPEAKERS	TOTAL NUMBER	PERCEPTUAL/ELIDED TAPS		MEASURABLE TAPS			
		N	%	N	%	DURATION	SD
Vf1	25	7	28	18	72	29	7
Vf2	17	13	76	4	24	30	10
Vf3	16	9	56	7	44	25	6
Vf4	27	14	52	13	48	21	8
Vf5	38	21	55	17	45	31	7
F total	123	64	52	59	48	27.2	7.6
Vm1	12	9	75	3	25	17	0
Vm2	38	14	37	24	63	29	7
Vm3	38	15	39	23	61	36	14
Vm4	2	1	50	1	50	34	0
Vm5	3	1	33	2	67	25	4
M total	93	40	43	53	57	28.2	5
M and F combined	216	104	48	112	52	27.7	10

Table 1. *Variation in the realization of phonological taps in VMS by speaker, type, and, for measurable taps, duration (in milliseconds).*

3.2. Trills

There were many allophonic variants of the trill in the corpus of VMS speech. We will first describe and illustrate the representative allophones and then present the data based on type grouping and by phonological conditions in order to identify patterns in VMS trill production. The data revealed a number of distinct realizations of the trill, as characterized by the narrowly transcribed phonetic categories in (2):

(2) a. [ɹ], [ɹ̥], [ɹ̝], [ɹ̝̊] — non-vibrant forms that include approximants and fricatives, both voiced and voiceless (figures 5 and 6)

b. [ɾɹ], [ɾɹ̝] — a single contact typically followed by vocalic *r*-coloring (figure 7) or frication (figure 8)

c. [rɹ], [r] — two or more contacts, both with and without additional vocalic *r*-coloring (figures 9 and 10)

Figures 5 and 6 give an example of voiced and voiceless fricative rhotics, respectively. Both realizations lack the brief contacts typically found in taps and trills, showing instead a longer duration of aperiodic noise concentrated around the 3000 Hz range. Following Solé (2002), we indicate the presence of fricative noise by adding a raising diacritic below the phonetic symbol for the approximant rhotic. A voice bar is present throughout the majority of the constriction period of the voiced [ɹ̝] but is absent during the voiceless [ɹ̝̊].

The example in figure 7 shows a single, brief contact followed by vocalic *r*-coloring, which is realized as continuous formant movement during the transition into the following vowel. In some cases, as illustrated in figure 8, the single contact is followed by a brief period of fricative noise combined with traces of formant structure.

In figure 9, two contacts are followed by an extensive period of *r*-coloring with visible formant movement into the following vowel. The example in figure 10 shows a trill consisting of three contacts, with brief intervening moments of periodic acoustic energy of higher intensity.

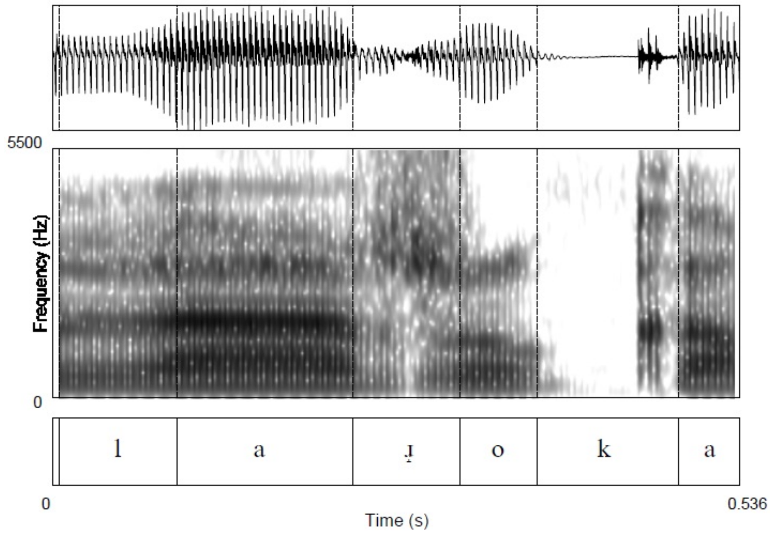

Figure 5. *Voiced fricative* [ɹ̝] *in VMS:* la roca.

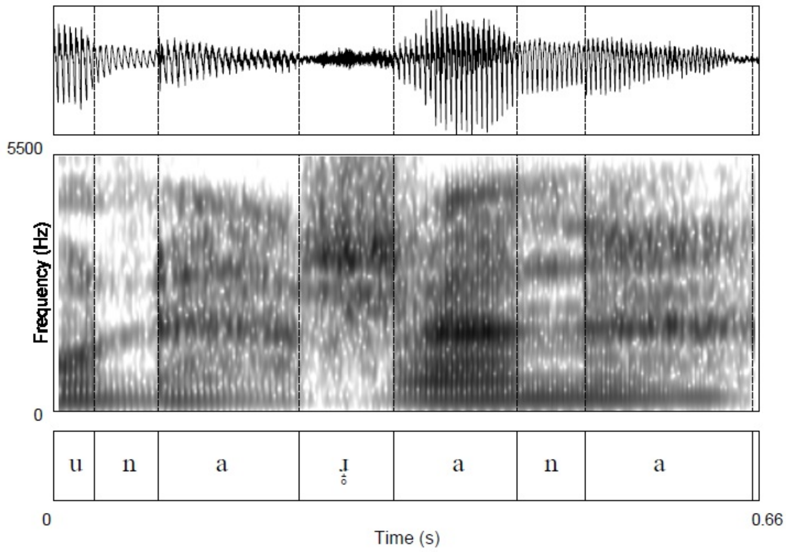

Figure 6. *Voiceless fricative* [ɹ̝̊] *in VMS:* una rana.

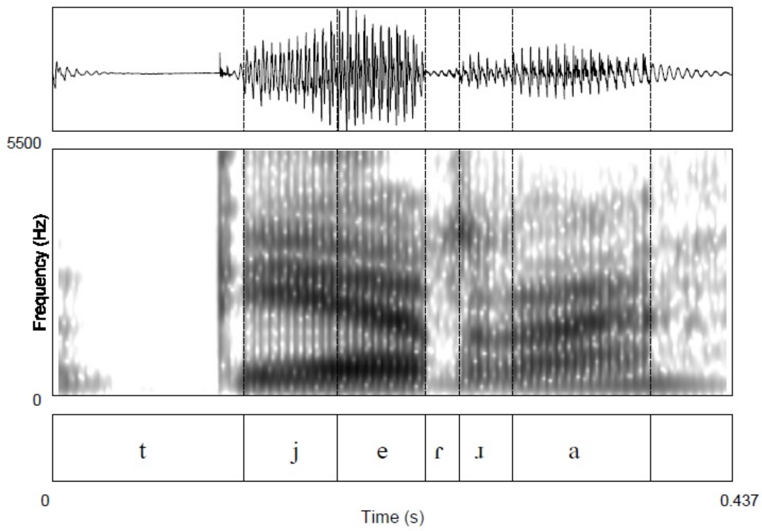

Figure 7. *Post-approximantized tap* [ɾɹ] *in VMS, with r-coloring:* perro.

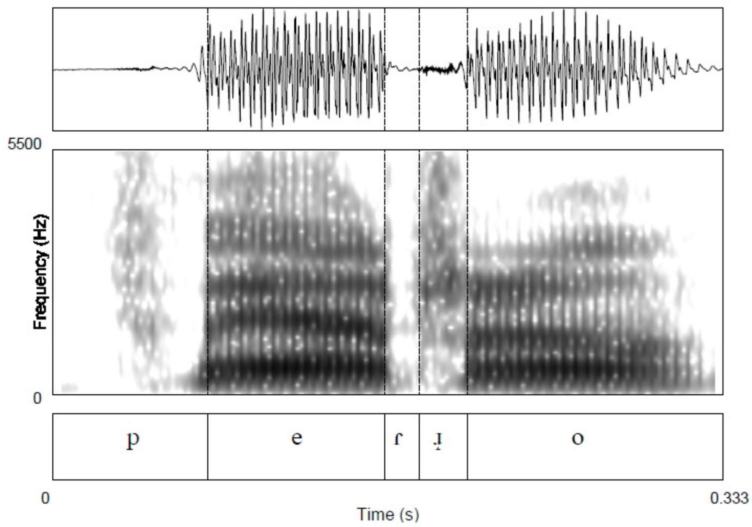

Figure 8. *Post-approximantized tap* [ɾɹ̝] *in VMS, with aperiodic noise:* perro.

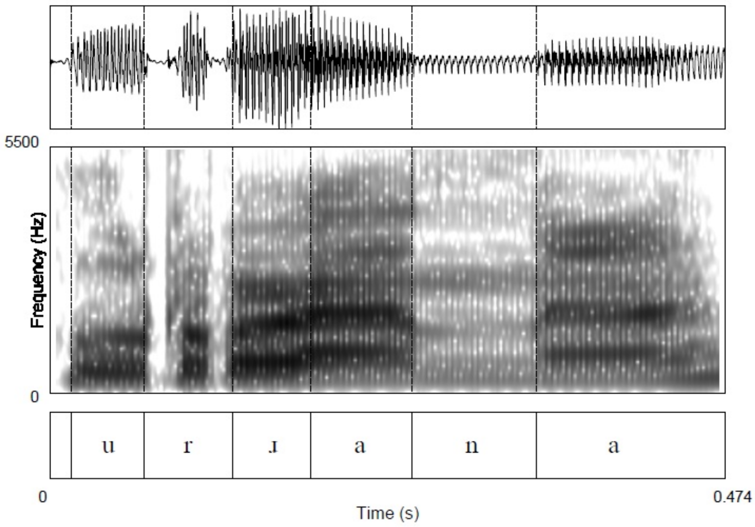

Figure 9. *Post-approximantized trill* [rɹ] *in VMS, with r-coloring:* (s)u rana.

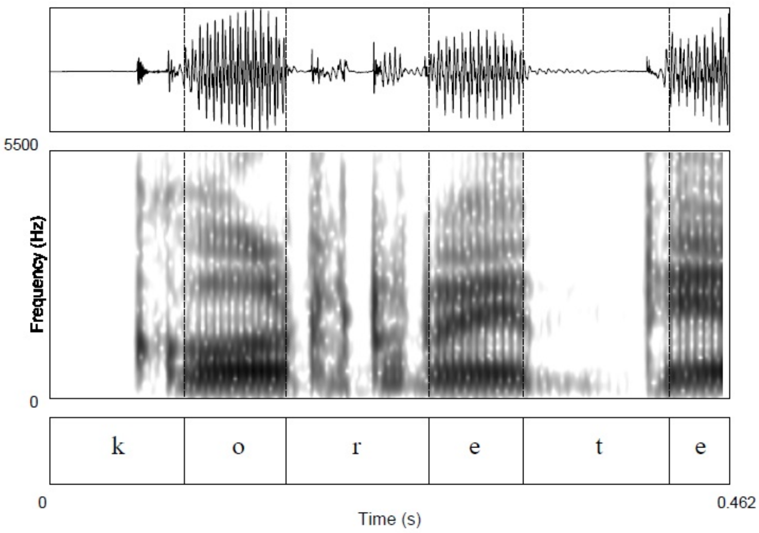

Figure 10. *Trill* [r] *in VMS, with three contacts:* correte(ando).

Productions similar to those in figures 7 through 9 have been documented in Peninsular Spanish and in other languages that include a trill in their phonological inventory. Lindau (1985: 160-161) describes a cross-linguistically common variant of the trill which lacks several pulses and instead *consists of a single closure followed by a prolonged opening phase. This variant occurs in all the languages that have trills.* Ladefoged and Maddieson (1996: 219-220) observe approximant phases that accompany lingual contacts in Italian trills in both word-initial and word-medial position. For Peninsular Spanish, Blecua (2001: 245-247) describes trill realizations consisting of one or two lingual contacts followed by an opening phase, and Martínez Celdrán and Fernández Planas (2007: 157-158) present spectrographic evidence of an approximant phase in the C-V transition between an intervocalic trill and the following vowel. For Russian and Peninsular Spanish trills, Kouznetsov and Pamies (2008, 2011) document realizations consisting of a single contact followed by a vocalic component before a following vowel. For Catalan, Recasens and Espinosa (2007: 13) report that trills in utterance-initial position *may show a burst followed by either a 20- to 40-ms steady-state approximant period immediately before the following vowel or a short frication noise.* Finally, Henriksen and Willis (2010) document trill productions in Jerezano Andalusian Spanish involving a single contact followed by *r*-coloring, as well as voiced and voiceless fricative variants.

For the purposes of the present study, we propose to unify the description of the VMS realizations in figures 7 through 9 as *post-approximantized* allophones of the trill. This label reflects the observation that continuous formant structure (with possible frication) is realized after the period of lingual contacts, which vary between one and two. Following Ladefoged and Maddieson (1996: 220), we indicate post-approximantization in narrow transcription by adding the symbol for the approximant [ɹ] (with or without the raising diacritic) after the tap or trill. While post-approximantized allophones are attested in Catalan, Italian, Peninsular Spanish, Russian, and other languages, to our knowledge this is the first time that such realizations have been empirically documented in a variety of Latin American Spanish.

As discussed in Section 1, trills in normative Spanish can be defined by the number of lingual contacts and by segmental duration. Researchers have reported variation in the number of contacts between two and five (Blecua, 2001; Recasens and Pallarès, 1999; Quilis, 1993), and average durations have been reported in the 60 ms range or higher (Blecua, 2001; Quilis, 1993). Following these descriptions, we define normative trills in the present study as having at least two visible contacts in the waveform and spectrogram, whereas non-normative realizations include non-vibrant forms and trills with one contact[5].

Table 2 provides a distribution of the trill tokens by speaker and by number of contacts. The most frequently occurring variants across the entire corpus are allophones with one or two contacts. Non-vibrant allophones and trills with three or more contacts are observed in much lower numbers. While recognizing considerable individual variation, we can note that most speakers produce some cases of all token types[6].

SPEAKERS	NON-VIBRANT	1 CONTACT	2 CONTACTS	3+ CONTACTS	TOTAL N
Vf1	1	22	12	4	39
Vf2	4	7	14	3	28
Vf3	3	2	14	17	36
Vf4	2	10	17	3	32
Vf5	0	19	19	6	44
Vm1	7	9	21	5	42
Vm2	11	23	7	0	41
Vm3	3	24	3	0	30
Vm4	5	11	6	1	23
Vm5	0	9	13	2	24
Totals	36	136	126	41	339

Table 2. *Variation in the realization of phonological trills in VMS by speaker and number of contacts.*

[5] Hammond (1999:142-143) considers all productions of the phonological trill with fewer than three contacts as non-normative according to his reading of norms established by the RAE. However, his categorization of trills was based primarily on auditory analysis alone.

[6] An anonymous reviewer suggests comparing the durations of categories to determine whether there are differences among the values associated with different types of trill realization, whether the differences are significant, and whether groups can be established. Such an investigation is beyond the scope of the present study, but we would expect overall duration to increase as the number of lingual contacts increases (see Blecua, 2001:§3.4.3.1 on Peninsular Spanish trills).

Table 3 highlights the variation in VMS trill production by collapsing the frequency data into non-normative and normative categories. All speakers produced both types of trill. The overall frequency of non-normative trill production by females was 39%, versus 64% for the male speakers. A chi-square test of independence was performed to examine the relation between sex and realization of phonological trills as non-normative versus normative. The relation between these variables was significant, χ^2 (1, N = 339) = 20.528, p < 0.01, suggesting that as a group, males were more likely to produce non-normative trills than were females, at least in the corpus produced by these particular VMS speakers.

SPEAKERS	TOTAL N	NON-NORMATIVE (0-1 CONTACTS)		NORMATIVE (2+ CONTACTS)	
		N	%	N	%
Vf1	39	23	59	16	41
Vf2	28	11	39	17	61
Vf3	36	5	14	31	86
Vf4	32	12	38	20	63
Vf5	44	19	43	25	57
F total	179	70	39	109	61
Vm1	42	16	38	26	62
Vm2	41	34	83	7	17
Vm3	30	27	90	3	10
Vm4	23	16	70	7	30
Vm5	24	9	38	15	63
M total	160	102	64	58	36
M and F combined	339	172	51	167	49

Table 3. *Variation in the realization of phonological trills in VMS by speaker and type (non-normative versus normative).*

In table 3, however, there was also considerable variation within each group of speakers. The frequency of non-normative trill production by females ranged from 14% to 59% versus 38% to 90% for the males. In particular, there was a noticeable tendency for three of the male speakers (Vm2, Vm3, and Vm4) and one female speaker (Vf1) to produce a higher percentage of non-normative trill forms. In the realization of VMS trills, it is possible that general inter-speaker variation is more of a relevant factor than sex.

Table 4 provides average duration measures of the VMS phonological trill as produced by each speaker, grouped by stress (i.e. trill allophones in the onset of tonic versus atonic syllables) and by word position (i.e. trill allophones in word-initial versus word-medial intervocalic position). Mann-Whitney's *U* tests were carried out using the mean durations as observations. The median duration values for trills were 80 ms in tonic versus 70 ms in atonic syllables, and 77 ms in initial versus 70 ms in word-medial positions. No significant differences were found with regard to stress ($U = 29.5$, $n_1 = n_2 = 10$, $p = 0.121$, two-tailed) or word-position ($U = 35$, $n_1 = n_2 = 10$, $p = 0.257$, two-tailed).

SPEAKERS	STRESS				WORD POSITION			
	TONIC		ATONIC		INITIAL		MEDIAL	
	N	DUR.(SD)	N	DUR.(SD)	N	DUR.(SD)	N	DUR.(SD)
Vf1	18	57 (22)	21	60 (25)	18	58 (22)	21	59 (25)
Vf2	19	79 (28)	9	76 (37)	24	78 (31)	4	79 (29)
Vf3	21	81 (22)	15	70 (26)	21	79 (22)	15	73 (27)
Vf4	22	65 (23)	10	57 (21)	25	62 (22)	7	66 (26)
Vf5	24	82 (33)	20	62 (20)	21	79 (30)	23	68 (27)
Vm1	18	115 (40)	24	85 (28)	25	108 (39)	17	83 (29)
Vm2	32	70 (18)	9	55 (18)	21	69 (19)	20	64 (20)
Vm3	14	68 (20)	16	78 (15)	13	71 (18)	17	70 (17)
Vm4	13	88 (31)	10	72 (33)	14	89 (24)	9	67 (40)
Vm5	9	81 (22)	15	70 (15)	13	76 (21)	11	73 (15)

Table 4. *Variation in the realization of phonological trills in VMS by speaker, stress, word position, and duration (in milliseconds).*

3.3. Tap/trill comparison

Comparisons based on number of lingual contacts and on average duration revealed appreciable differences between the two rhotics in VMS, including allophonic variation, post-approximantization with or without aperiodic noise, and mostly non-overlapping segmental duration. Table 5 compares the average durations of phonological tap and trill realizations in the contrastive, word-medial intervocalic context. Although low token numbers require caution in the interpretation of these results, some trends can still be observed. Differences in standard deviation indicate that trills generally showed a greater range of variation in segmental duration as compared to taps. The tap duration averages, which ranged from 17 ms for speaker Vm1 to 36 ms for speaker Vm3, contrast with those of the trill allophones, which were in or above the 60 ms range for the majority of speakers. The tap/trill duration distributions were partially overlapping only for speaker Vf1. Taking the mean durations as observations, a Mann-Whitney's *U* test indicated that median trill duration, 69 ms, was greater than median tap duration, 29 ms ($U = 100$, $n_1 = n_2 = 10$, $p < 0.001$, one-tailed). This suggests that intervocalic trills were significantly longer than intervocalic taps, at least in the corpus produced by these particular speakers of VMS.

SPEAKERS	TAP			TRILL		
	N	DURATION	SD	N	DURATION	SD
Vf1	18	29	7	21	59	25
Vf2	4	30	10	4	79	29
Vf3	7	25	6	15	73	27
Vf4	13	21	8	7	66	26
Vf5	17	31	7	23	68	27
Vm1	3	17	0	17	83	29
Vm2	24	29	7	20	64	20
Vm3	23	36	14	17	70	17
Vm4	1	34	0	9	67	40
Vm5	2	25	4	11	73	15

Table 5. *Average duration (in milliseconds) of phonological taps and trills in word-medial intervocalic position.*

It was pointed out in the introduction that the phonetic realization of syllable-initial trills serves to distinguish among different varieties of Spanish and that experimental studies have documented a wide range of trill variation in terms of manner, place, and voicing. In particular, Hammond (1999, 2000a,b, 2006) finds little evidence of trills realized with three lingual contacts in normal discourse in the speech of the vast majority of native speakers, and he claims that the tap/trill contrast has been neutralized in many dialects. With respect to the VMS data, the claim of contrast neutralization might be supported if the comparison were based exclusively on the number of contacts and if only allophones with three contacts were counted as successful realizations of the phonological trill. As shown in table 2, only 41 (i.e. 12%) of 339 total trills were realized with three or more contacts, while 298 (i.e. 88%) were realized with two or fewer contacts. However, as illustrated in table 5, the durations of phonological taps and trills were significantly different, with trills generally two to three times as long as taps[7]. Thus, while intervocalic trills in the VMS corpus showed a reduction in the degree of lingual trilling, acoustic analysis revealed a difference in overall duration as compared to taps. To be sure, it remains to be determined whether the observed durational differences between allophones of the VMS tap and trill are actually recoverable by listeners from this and other dialects of Spanish. We leave this task to future research, which should address the perceptual robustness of the segmental contrast both within VMS as well as for non-VMS speakers.

4. DISCUSSION

In this section, we relate our findings to recent experimental work on rhotic variation in Dominican Spanish. We then offer a phonetic explanation of the patterning of syllable-initial rhotics in both Spanish varieties based on articulatory and aerodynamic factors.

4.1. Cross-dialectal comparison with rhotic variation in Dominican Spanish

Dominican Spanish (henceforth, DS) has a widely recognized modification of the trill, which has traditionally been described in terms of *pre-aspiration* (Jiménez

[7] These numbers reflect the measurable tap tokens only (112 or 52% of the tap corpus). The other 48% were reduced or elided forms without consistent landmarks for measurements and, if included, would reflect a zero measurement resulting in a skewed portrayal of tap duration.

Sabater, 1975; Lipski, 1994; Núñez Cedeño, 1989, 1994). However, recent laboratory studies have called into question the traditional description. On the basis of acoustic data collected in a story narration task with speakers from Santo Domingo and the Cibao Valley, Willis (2006, 2007) argues that *pre-aspiration* is a misnomer and that the DS trill is best characterized as a period of *pre-breathy voice* followed by one or more contacts.

Figures 11 and 12 show examples of the two most frequently occurring allophones of the DS trill, transcribed as [ɦɾ] and [ɦr]. Both the breathiness and the lingual contacts of these allophones are typically voiced, in contrast to previous, impressionistic descriptions of these components as voiceless. Willis (2006, 2007) finds that the two voiced allophones together account for approximately 75% of the total productions in the DS corpus. Furthermore, variation in trill production is not limited to the contrastive intervocalic position, as pre-breathy voiced allophones are found both word-medially (figure 11) and word-initially (figure 12). The same is true of post-approximantized allophones of the VMS trill (see figures 7 through 9). This suggests that in contemporary Spanish, variation affects syllable-initial trills more generally and that non-normative allophones of the trill can appear word-initially as well as word-medially.

In an acoustic comparison of DS taps and trills, Willis and Bradley (2008) document the same range of allophonic realizations of intervocalic /ɾ/ in DS as observed in VMS, as well as comparable rates of tap reduction[8]. Out of a total of 154 phonological taps in the DS corpus, 51% (79) were produced with a measurable contact (combining the noncontinuant and approximant allophones), while the remaining 49% (75) were produced without a measurable contact.

[8] One difference, however, is that in the DS corpus, noncontinuant tap productions do not have a release burst as strong as the one sometimes observed on VMS taps (see Section 3.1).

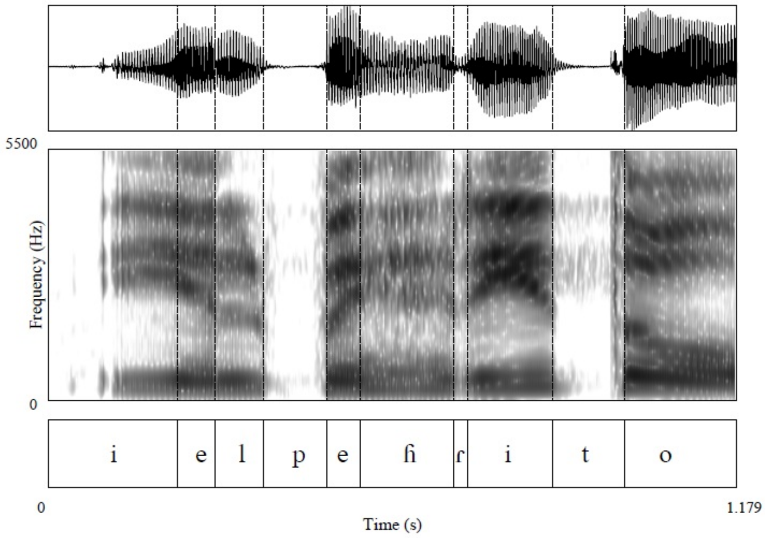

Figure 11. *Pre-breathy-voiced tap* [ɦɾ] *in DS, with one contact:* y el perrito.

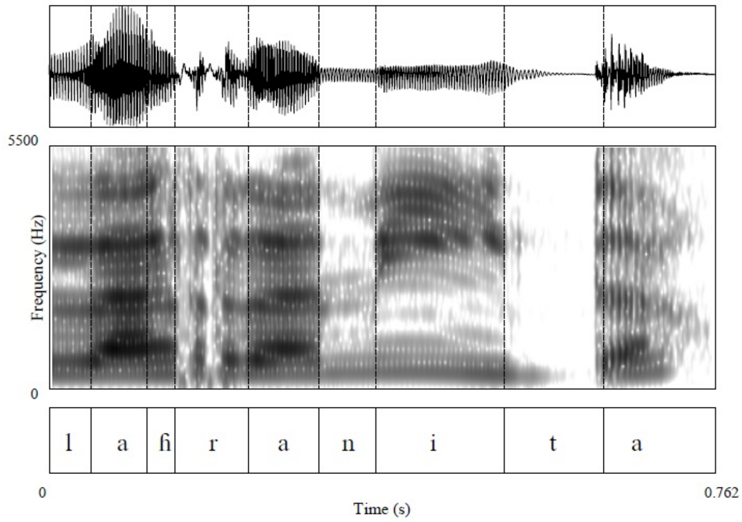

Figure 12. *Pre-breathy-voiced trill* [ɦr] *in DS, with three contacts:* la ranita.

Given that the DS trill is commonly realized with a single contact, as illustrated in figure 11, the claim of rhotic neutralization might be supported if the comparison is based exclusively on the number of contacts. However, Willis and Bradley (2008) report that the overall duration of the two contrastive segments is notably different, with the average duration of trill allophones typically three times as long as the average duration of taps. Across speakers and dialect regions, average tap duration in DS ranges from 20-23 ms, whereas average trill duration ranges from 72-89 ms. The averages and standard deviations of the two rhotic categories are non-overlapping within each group of speakers. Thus, while the DS trill does tend to have a reduced number of contacts compared to normative Spanish, the overall segmental duration reveals a clear acoustic cue for contrast maintenance.

Taken together with the findings from VMS (see Section 3.3), the DS data suggest that the contrast between the trill and the tap can be maintained in terms of overall segmental duration even when there is no difference in the number of lingual contacts. These results are not surprising, given the range of phonetic variation in trills attested across Spanish dialects. While neutralization is attested in Judeo-Spanish dialects that have lost the trill in favor of the tap (Quintana, 2006: 84-85, 376), it is clear that the contrast is still robustly manifested in both VMS and DS. Therefore, claims of rhotic neutralization should be reserved for varieties like those of Judeo-Spanish in which /ɾ/ is the only remaining rhotic in the phonological inventory.

4.2. Phonetic explanation of syllable-initial rhotic variation

The cross-dialectal comparison suggests that pre-breathy voicing in DS and post-approximantization in VMS can be understood as alternate strategies for articulating the voiced multiple-contract trill. In an experiment investigating the relationship among oropharyngeal pressure, airflow, and lingual trilling, Solé (2002) used different sized catheters to vent pressure from behind the back molars while subjects pronounced voiced and voiceless trills. Results indicate that voiced trills place a high degree of aerodynamic constraint on oropharyngeal pressure in order to maintain simultaneous lingual and glottal vibration. Small changes in pressure may result in the loss of trilling, voicing, or both.

Willis and Bradley (2008) argue that the innovation of breathy voicing in DS trills can be viewed as an articulatory strategy to create favorable aerodynamic conditions for the initiation of lingual trilling. Solé (2002: 675) observes that a higher pressure difference is needed to initiate trilling than to sustain it, and this explains why trills are *pre-* and not *post*-breathy voiced. As pre-breathy voiced trills began to appear more frequently in the speech of DS speakers, listeners may have reinterpreted the breathy voice component as a formal part of the segment's representation. With pre-breathy voicing as a compensatory durational cue, the lingual trilling gesture could then lenite to a pre-breathy voiced tap without neutralizing the rhotic contrast (Zlotchew, 1974). Willis and Bradley (2008: 97-98) suggest that DS trill allophones may be situated along a continuum of progressively less effortful forms, [r] > [ɦr] > [ɦɾ] > [ɦ], and they provide gestural scores illustrating the articulatory changes involved[9]. The idea of connecting rhotic realizations along lenition scales is not new (see Blecua, 2001: §§4.2-4.5)

We would like to argue that post-approximantization in VMS also has an aerodynamic basis. As noted above, Solé (2002) finds that a higher pressure difference is needed to initiate trilling than to sustain it. Many of the trills produced in her experimental study often decayed into fricatives or approximants when oropharyngeal and subglottal pressure fell below a certain threshold. Ladefoged and Maddieson (1996: 219) explain the approximant phases observed in Italian trills as the result of the tongue not being held close enough to the upper surface of the mouth for sustained trilling. Recasens and Espinosa (2007) adopt the same explanation for Catalan. We propose that the post-approximantized allophones in VMS result from a successful initiation of lingual trilling followed by its extinction into an approximant, which shows up as *r*-coloring in the transition from the rhotic to the following vowel. As argued in Section 3.3, when trills are realized with a single contact, the additional acoustic component increases the total duration of the rhotic segment. Those taps produced with a measurable contact are short enough to maintain an acoustic contrast with the post-approximantized allophones, which can be situated along a continuum of trill lenition: [rː] > [rɹ] > [ɾɹ], where [rː] and [r] denote three and two lingual contacts, respectively.

Our view of post-approximantization is consistent with reports by Kouznetsov and Pamies (2008, 2011) that trills in Russian and Peninsular Spanish can be produced

[9] Perceptual factors may also be involved in the innovation of DS pre-breathy voiced rhotics. Although they are more aerodynamically stable than voiced trills, voiceless trills are perceptually very similar to fricatives (Solé, 2002: 680-682), which increases the potential confusability of the two segment types. Breathy voicing can be seen as a compromise between modally voiced and voiceless trills, combining the perceptual distinctiveness of the former and the aerodynamic stability of the latter. In addition, pre-breathy voice timing may be favored because it helps maintain clearer C-V transitions. See Willis and Bradley (2008: 95-98) for further discussion.

with only one closure followed by a vocalic component before the following vowel. In their Spanish data, prevocalic taps never show this vocalic component, and we have found that the same is true of taps in both VMS and DS. The generalization emerging here is that number of lingual contacts is not a reliable phonetic cue to the phonological contrast between rhotics, which can instead be expressed by other means, such as overall segment duration.

The cross-dialectal comparison reveals similar patterns of reduction and elision of intervocalic taps in both DS and VMS. The allophonic realizations of taps may also be seen as points along a lenition continuum, possibly as a function of articulatory effort. The sequence of phonetic forms [ɾ] > [ɾ̞] > [$^{(ɾ)}$] > [Ø] would constitute progressively more lenited realizations, with the final three forms showing greater degrees of reduction in lingual contact.

The idea that rhotic realizations are situated along a lenition scale in relation to interacting articulatory and perceptual factors is developed in greater detail by Blecua (2001: §§4.2-4.5). Her phonetically-based account of rhotic variation encompasses both taps and trills in Peninsular Spanish and includes different syllabic contexts and speech styles. Recent studies of VMS and DS rhotics lend further support to a phonetically-based explanation of trill variation by showing that post-approximantization exists in Spanish beyond the Peninsular variety and by documenting the existence of pre-breathy voiced rhotics, which to our knowledge have never been described for Peninsular Spanish. Our understanding of cross-linguistic variation in rhotics grows as insights from phonetic studies are accumulated, which invites further research on both Spanish and other languages.

ACKNOWLEDGMENTS: This work was first presented at the 38th Linguistic Symposium on Romance Languages, held in April, 2008 at the University of Illinois, Urbana-Champaign. We are grateful to audience members for feedback and criticism, especially Stuart Davis, Haike Jacobs, Rafael Núñez Cedeño, and Miquel Simonet. Thanks also to four anonymous reviewers of earlier versions of this work. We are responsible for any remaining shortcomings in the present article.

5. BIBLIOGRAPHIC REFERENCES

BLECUA, B. (2001): *Las vibrantes del español: manifestaciones acústicas y procesos fonéticos,* Ph. D. dissertation, Universitat Autònoma de Barcelona.

BOERSMA, P. and D. WEENINK (2001): «Praat, a system for doing phonetics by computer», *Glot International,* 5, pp. 341-345.

BRADLEY, T. G. (2006): «Phonetic realizations of /sr/ clusters in Latin American Spanish», in M. Díaz-Campos (ed.): *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology,* Somerville, MA, Cascadilla Proceedings Project, pp. 1-13.

COLANTONI, L. (2001): *Mergers, chain shifts and dissimilatory processes: palatals and rhotics in Argentine Spanish,* Ph. D. dissertation, University of Minnesota.

COLANTONI, L. (2006a): «Increasing periodicity to reduce similarity: an acoustic account of deassibilation in rhotics», in M. Díaz-Campos (ed.): *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology,* Somerville, MA, Cascadilla Proceedings Project, pp. 22-34.

COLANTONI, L. (2006b): «Micro and macro sound variation and change in Argentine Spanish», in N. Sagarra and A. J. Toribio (eds.): *Selected Proceedings of the 9th Hispanic Linguistics Symposium,* Somerville, MA, Cascadilla Proceedings Project, pp. 91-102.

DÍAZ-CAMPOS, M. (2008): «Variable production of the trill in spontaneous speech: sociolinguistic implications», in L. Colantoni and J. Steele (eds.): *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology,* Somerville, MA, Cascadilla Proceedings Project, pp. 47-58.

HAMMOND, R. (1999): «On the non-occurrence of the phone [r̃] in the Spanish sound system», in J. Gutiérrez-Rexach and F. Martínez-Gil (eds.): *Advances in Hispanic Linguistics,* Somerville, MA, Cascadilla Press, pp. 135-151.

HAMMOND, R. (2000a): «The multiple vibrant liquid in the discourse of U.S. Hispanics», in A. Roca (ed.): *Spanish in the US: Theoretical and Applied Perspectives,* Somerville, MA, Cascadilla Press, pp. 290-304.

HAMMOND, R. (2000b): «The phonetic realizations of /rr/ in Spanish: a psychoacoustic analysis», in H. Campos, E. Herburger, A. Morales-Front, and T. J. Walsh (eds.): *Hispanic Linguistics at the Turn of the Millennium,* Somerville, MA, Cascadilla Press, pp. 80-100.

HAMMOND, R. (2006): «The status of [ɾ] and [r̃] in Spanish: A functional analysis», in M. Sedano, A. Bolívar, and M. Shiro (eds.): *Haciendo lingüística: Homenaje a Paola Bentivoglio,* Caracas, Universidad Central de Venezuela, pp. 91-104.

HENRIKSEN, N. and E. WILLIS (2010): «Acoustic characterization of phonemic trill production in Jerezano Andalusian Spanish», in M. Ortega-Llebaria (ed.): *Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology,* Somerville, MA, Cascadilla Proceedings Project, pp. 115-127.

HUALDE, J. I. (2005): *The Sounds of Spanish,* Cambridge, Cambridge University Press.

JIMÉNEZ SABATER, M. (1975): *Más datos sobre el español en la República Dominicana,* Santo Domingo, Ediciones Intec.

KOCHETOV, A. and L. COLANTONI (2011): «Coronal place contrasts in Argentine and Cuban Spanish: An electropalatographic study», *Journal of the International Phonetic Association,* 41, pp. 313-342.

KOUZNETSOV, V. and A. PAMIES (2008): «Trill with one closure, still a trill or a tap? Data from Russian and Spanish», in *Proceedings of the XX Session of the Russian Acoustical Society,* Moscow, Russian Acoustical Society, pp. 672-675.

KOUZNETSOV, V. and A. PAMIES (2011): «Enfoque perceptivo y contrastivo de la vibrante múltiple», Paper presented at the *V Congreso Internacional de Fonética Experimental,* University of Extramadura.

LADEFOGED, P. and I. MADDIESON (1996): *The Sounds of the World's Languages,* Oxford, Blackwell.

LEWIS, A. (2004): «Coarticulatory effects on Spanish trill production», in A. Agwuele, W. Warren, and S. H. Park (eds.): *Proceedings of the 2003 Texas Linguistics Society Conference,* Somerville, MA, Cascadilla Proceedings Project, pp. 116-127.

LINDAU, M. (1985): «The story of /r/», in V. Fromkin (ed.): *Phonetic Linguistics. Essays in Honour of Peter Ladefoged,* pp. 157-168.

LIPSKI, J. (1994): *Latin American Spanish,* New York, Longman.

MARTÍNEZ CELDRÁN, E. and A. M. FERNÁNDEZ PLANAS (2007): *Manual de fonética española: articulaciones y sonidos del español,* Barcelona, Ariel.

MAYER, M. (1969): *Frog, Where Are You?,* New York, Dial Press.

NÚÑEZ CEDEÑO, R. (1989): «La /r/, único fonema vibrante del español: datos del Caribe», *Anuario de Lingüística Hispánica,* 5, pp. 153-171.

NÚÑEZ CEDEÑO, R. (1994): «The alterability of Spanish geminates and its effects on the Uniform Applicability Condition», *Probus,* 6, pp. 23-41.

QUILIS, A. (1993): *Tratado de fonología y fonética españolas,* Madrid, Arco.

QUINTANA, A. (2006): *Geografía lingüística del judeoespañol: estudio sincrónico y diacrónico,* Bern, Peter Lang.

RECASENS, D. (1991): «On the production characteristics of apicoalveolar taps and trills», *Journal of Phonetics,* 19, pp. 267-280.

RECASENS, D. and A. ESPINOSA (2007): «Phonetic typology and positional allophones for alveolar rhotics in Catalan», *Phonetica,* 64, pp. 1-28.

RECASENS, D. and M. D. PALLARÈS (1999): «A study of /r/ and /rr/ in the light of the 'DAC' coarticulation model», *Journal of Phonetics,* 27, pp. 143-169.

SESSAREGO, S. (2011): «Phonetic analysis of /sr/ clusters in Cochabambino Spanish», in L. A. Ortiz-López (ed.): *Selected Proceedings of the 13th Hispanic Linguistics Symposium,* Somerville, MA, Cascadilla Proceedings Project, pp. 251-263.

SIMONET, M. and P. CARRASCO (2006): «Acoustic profiling of word-initial rhotics in Costa Rican Spanish», Paper presented at the *3rd Conference on Laboratory Approaches to Spanish Phonology,* University of Toronto.

SOLÉ, M.J. (2002): «Aerodynamic characteristics of trills and phonological patterning», *Journal of Phonetics,* 30, pp. 655-688.

WILLIS, E. (2006): «Trill variation in Dominican Spanish: an acoustic examination and comparative analysis», in N. Sagarra and A. J. Toribio (eds.): *Selected Proceedings of the 9th Hispanic Linguistics Symposium,* Somerville, MA, Cascadilla Proceedings Project, pp. 121-131.

WILLIS, E. (2007): «An acoustic study of the "pre-aspirated trill" in narrative Cibaeño Dominican Spanish», *Journal of the International Phonetic Association,* 37, pp. 33-49.

WILLIS, E. and T. G. BRADLEY (2008): «Contrast maintenance of taps and trills in Dominican Spanish: data and analysis», in L. Colantoni and J. Steele (eds.): *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology,* Somerville, MA, Cascadilla Proceedings Project, pp. 87-100.

ZAMORA, J. and J. GUITART (1988): *Dialectología hispanoamericana: teoría, descripción, historia,* Salamanca, Almar.

ZLOTCHEW, C. (1974): «The transformation of the multiple vibrant to the fricative velar in the Spanish of Puerto Rico», *Orbis,* 23, pp. 81-84.