Gestural Timing in the Perception of Spanish r+C Clusters

Joaquín Romero
Universitat Rovira i Virgili

1. Introduction

Spanish single <r> (*erre simple*) occurs in complementary distribution with double <r> (*erre doble*) and appears everywhere except in absolute word-initial position and following /n/, /s/ and /l/ (Navarro Tomás 1970). According to Navarro Tomás the articulation of single <r> (henceforth /ɾ/) consists of a quick elevation of the tongue tip towards the alveolar region followed by a weak and brief contact and a subsequent return to the starting position. This description would correspond to that of a 'tap' (see also Hualde 2005). Articulatory evidence (Romero 1996), however, shows that /ɾ/ can also be produced as a 'flap', in which case the tongue tip follows a more complex trajectory, moving upwards and backwards slightly before flapping forward, hitting the alveolar ridge in the process and then returning to the resting position. This would be the situation, for example, in intervocalic position, as in *para* /ˈpaɾa/ 'for'. In other positions, such as preceding or following consonants, single <r> is accompanied by a well-known vocalic element, commonly referred to as svarabhakti vowel (Gili Gaya 1921, Malmberg 1965, Martínez Celdrán 1978). According to Quilis (1970), this vocalic element can be of equal or even greater length than the <r> itself and is usually a copy in quality of the flanking vowel, so that a word like *crónica* 'chronicle' would actually be pronounced as something nearing *corónica*.

The exact nature of this vocalic element is examined in this paper from an articulatory as well as a perceptual point of view. The main hypothesis that is put forth here is that the vocalic element is not the result of any independent process of epenthesis, but rather comes as a natural consequence of the articulatory nature of /ɾ/. This hypothesis is tested by investigating the articulatory nature of /ɾ/ vis-à-vis /s/, in particular whether these two tongue-tip consonants can be minimally distinguished perceptually on the basis of duration (Romero 2001, Romero & Martín 2003). A perceptual experiment is presented which investigates the relevance of the presence versus absence of the vocalic element, as well as its length, to the perception of /ɾ/ vis-à-vis /s/. The results suggest that the so-called svarabhakti vowel is not absolutely necessary for the perception of r+C clusters, even though its presence increases the rate of correct identification. Some conclusions are also presented concerning the role of duration in the phonological description of Spanish single <r>.

1.1. Articulatory characteristics of /ɾ/

Romero (1996) presents electromagnetometer data on the production of Spanish /ɾ/ in a variety of contexts, including intervocalic position (as in the sequences /eɾe/ and /oɾo/) and in preconsonantal position (including /eɾde/, /oɾdo/, /eɾge/ and /oɾgo/). Figure 1 reproduces some of the tongue-tip movement data in Romero (1996) in the context of and in comparison with other tongue-tip consonants. The lines in this figure represent the movement in two-dimensional space of the tongue tip as traced by the tongue-tip coil during the production of VC(C)V sequences (see Romero 1995 for details on electromagnetometer data collection and display).

As can be seen in the figure, whether in intervocalic position (single /ɾ/ in the graphs) or in preconsonantal position (/ɾd/ sequences), the tongue tip trajectory corresponds to the flapping movement mentioned earlier. This movement is more extreme in the context of the back vowel /o/ because the tongue tip trajectory for /ɾ/ starts from a lower, more retracted position than with the front

 Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology, ed. Laura Colantoni and Jeffrey Steele, 59-71. Somerville, MA: Cascadilla Proceedings Project.

vowel /e/, but even with /e/ the shape of the trajectory is very similar and clearly indicative of the presence of a flap.

As far as the /ɾd/ sequences are concerned, the paths in Figure 1 show patterns of movement that resemble those of single /ɾ/, with the forward sliding movement of the tongue tip starting in approximately the same area and ending in a very similar position as for single intervocalic /d/. This is taken as evidence that the /ɾ/ and /d/ are articulated in sequence and that there is no blending of the constriction locations of the two gestures.

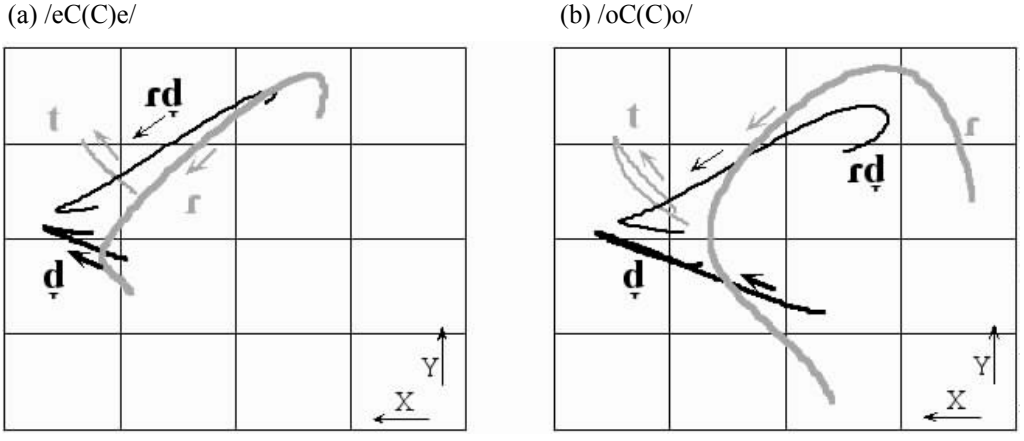

Figure 1. Two-dimensional EMMA tongue-tip movement paths for intervocalic /ɾ/ (thick grey line), /t/ (thin grey line), /d/ (thick black line) and /ɾd/ (thin black line) in the context of the vowel /e/ (a) and in the context of the vowel /o/ (b). Arrows next to the paths indicate the direction of the movement of the tongue tip. The diacritics in /d̬/ and /ɾd̬/ represent the non-occlusive realization of this consonant.

This sequential articulation is interpreted by Bradley (2004) in the framework of Articulatory Phonology (Browman & Goldstein 1986, 1989, 1992a) as a case of lack of articulatory overlap in the timing between the adjacent tongue tip gestures for the /ɾ/ and /d/. The separation between these two gestures results in the perception of a small portion of the flanking vowel, which would correspond to the above-mentioned svarabhakti vowel. From the trajectories in Figure 1, the portion of the forward sliding movement between the starting position and the end point corresponding to the constriction location for /d/ would correspond to the time period during which the flanking vowel is audible. Clearly this explanation hinges on the assumption that /ɾ/ is articulated as a flap, for the default tap realization assumed by Hualde (2005) and others would involve the simple elevation of the tongue tip towards the alveolar ridge, therefore lacking the sliding movement necessary for the perception of the vocalic portion. Such a tapped articulation would in fact resemble that of an alveolar lateral /l/. The fact that there is no vocalic element in /ld/ sequences, or even deocclusivization of the stop, is a good indication that the tongue tip gestures for /ɾ/ and /l/ are different.

In his study, Bradley (2004) also shows cases of the so-called assibilated single <r> from highland Ecuadorian Spanish, that is cases where /ɾ/ is pronounced as a fricative segment, commonly represented as /ř/. This assibilated <r> is widespread in many parts of the Spanish-speaking world (Argüello 1978, Hualde 2005) and can even be heard in weak productions of trills in Castilian and other dialects of Spanish as well as in Catalan (Recasens submitted) and languages like Swedish (Granberry 1999). Bradley proposes a gestural explanation in order to distinguish between assibilated <r> and regular /ɾ/ in /ɾt/ sequences on the basis of articulatory timing: whereas in /ɾt/ sequences there is no overlap between the adjacent tongue-tip gestures, which results in the perceived vocalic element, in the assibilation case, increased overlap between the gestures for /ɾ/ and /t/ prevents the perception of the vocalic element. In addition, increased duration results in the formation of a longer constriction which yields the fricative or assibilated consonant.

Whereas this explanation seems reasonable from an articulatory or gestural point of view, at least as far as timing is concerned, what is missing from Bradley's explanation is the fact that the actual

tongue-tip gestures for the non-overlapped, unblended /ɾt/ and for the overlapped, blended /ř̥t/ are probably different to begin with. If the difference between the two conditions were exclusively a matter of timing, then the effect of increased overlap would likely be a shorter duration of /ɾ/, resulting in a blended sequence like the ones that are common in some dialects of Caribbean Spanish (Guitart 1976), where /ɾt/ sequences often result in a somewhat retroflexed long or geminated consonant /ʈ̚ʈ/. The retroflexion would seem reasonable since the shortening of /ɾ/ would basically reduce it to the initial part of the flapping movement. The fact that there is no retroflexion in assibilated /ř̥/, at least as far as Bradley's observations are concerned, probably indicates that there is no flapping movement in these cases to begin with. Instead, the tongue tip gesture is probably more tap-like in these cases, lacking the forward sliding movement of the flap, which would result in an apical fricative given the appropriate timing and aerodynamic conditions.

1.2. Vowel epenthesis versus gestural coordination

Independent of the specifics of each particular case, an explanation of these types of phenomena along gestural lines seems more faithful to reality, at least from an articulatory point of view, than those that have interpreted the vocalic element in Spanish r+C clusters as a case of simple vowel epenthesis (Ravelo 1999). Other comparable phenomena which have been analyzed as instances of vowel epenthesis can in fact be understood in similar gestural terms. Colantoni & Steele (2005) show evidence of vocalic epenthesis in obstruent-liquid clusters in French and Spanish. While the authors recognize the phonetic nature of the process, they do not explicitly contemplate the possibility of an articulatory explanation in terms of gestural overlap. Another such case is the often-called epenthetic schwa that is heard in vowel+/l/ and vowel+/r/ sequences in English (Olive, Greenwood & Coleman 1993, Wells 2000), as in *feel* or *fear* transcribed as [fiˑl] and [fiˑɹ]. Riera & Romero (2006) provide preliminary acoustic data that indicate that this so-called schwa is in fact not the result of an independently motivated process of vowel epenthesis, but rather follows naturally from a situation of articulatory overlap between the tongue-tip and tongue-body gestures in the coda vowel-consonant sequences, as supported by a high level of spectral variability across different vocalic contexts. The fact that both the English schwa and the Spanish svarabhakti vowels are also highly variable in duration (Ramírez 2006, Schmeiser 2006) is another good indication that what we are dealing with here is not the expected contextual variability of an independently motivated phonological process of vowel epenthesis or insertion, but rather the contextually conditioned articulatory transitions between vowel and consonant, in the English case, and between consonant and consonant, in the Spanish case. In other words, both the English schwa and the Spanish svarabhakti vowels would be epiphenomena, by-products of the temporal overlap between articulatory gestures, not the result of a deliberate process of insertion.

Even though these two views result from conceptually different ways of understanding phonology and its relationship with speech production, the reality is that it is often difficult if not impossible to determine whether a particular phenomenon is 'automatic' or 'controlled' (Fowler 1990, Kingston & Diehl 1994). In the case of the Spanish svarabhakti vowels we have assumed that a high degree of variability in the nature of the vowels as well as in their duration can be interpreted as pointing to a 'targetless' segment (Browman & Goldstein 1992b), that is, a segment with no specifications in the phonological planning, which would support an 'automatic' interpretation of the phenomenon. One possible way to test the validity of this assumption is to study how the variability observed in the production affects the perception. Though the relationship between speech production and perception continues to be a source of heated debate (Liberman & Mattingly 1985, Fowler 1996, Lindblom 1996), it would be interesting to find out simply to what extent listeners rely on the presence and/or characteristics of the svarabhakti vowels in terms of duration in order to correctly identify the r+C sequences.

With these issues in mind, the present study attempts to test the perceptual relevance of the vocalic element in Spanish /ɾ/+C clusters. If, as predicted, the vocalic element is the result of gestural coordination and not a deliberate epenthetic segment, then we might conclude that its presence or absence should not be crucial for the perception of /ɾ/+C clusters. In other words, speakers should not

be expected to rely on the presence of an epiphenomenal vowel in order to identify the /ɾ/+C sequence. Instead they should be able to identify such sequences simply from the presence of the appropriate starting and end points in the /ɾ/ flapping gesture. Other related questions that are addressed in this study concern the duration of the vocalic element. In the event that the vocalic element turns out to be crucial for the perception of /ɾ/+C clusters, it would be interesting to determine whether a reduction in duration decreases significantly the correct identification of these sequences. This would also have implications as far the duration of /ɾ/ itself is concerned, as well as to whether the exact shape of the tongue-tip gesture for /ɾ/, (i.e., more flap-like or tap-like), and its exact constriction location are important in the perception.

2. Method

2.1. Previous findings

The experimental methodology used in this study builds on earlier studies (Romero 2001, Romero & Martín 2003) that investigated the relevance of duration in the perception of reduced tongue-tip consonants cross-linguistically from a historical point of view. The working hypothesis in those studies was based on the notion that the historical sound change known as rhotacism (i.e., in segmental terms a change from an /s/-type consonant to a rhotic, usually a tap), could be understood in a more straightforward manner by looking at the process as basically a reduction in the duration of the original fricative, which could result in a gesture that resembles, and eventually can be identified as, a rhotic. Romero (2001) investigated the /z/ to /ɾ/ change in syllable onset position in Catalan, while Romero & Martín (2003) looked at the /t/ to /ɾ/ process also in syllable onsets in American English. The methodology employed in these studies consisted in reducing the duration of an original /z/ to match that of an original /ɾ/[1] in order to test whether a simple reduction in the duration of the original /z/ could induce the percept of /ɾ/. In both cases the results were clear in that the percept of /ɾ/ was categorically achieved simply by shortening the duration of /z/ beyond a certain point that approximated the duration of the original /ɾ/. These results were taken as evidence that historical rhotacism is more adequately explained in terms of reduction in the magnitude of the tongue-tip gesture than as a case of segmental substitution.

In the case of Spanish, the observation that synchronic rhotacism of syllable-final /s/ is a spreading phenomenon, at least in Castilian dialects, resulted in a slightly more complex situation than that of Catalan or English, precisely because of the presence of a transitional vocalic element between the /ɾ/ and the following consonant, which never arises in syllable onset position. Colloquial Castilian Spanish, particularly varieties of the Madrid and Northern Castile areas, often rhoticizes syllable-final /s/ in front of continuants including [b̥], [d̥] and /θ/, as in *dos veces* [doɾ b̥eθes] ‘two times’, *desde* [deɾd̥e] ‘since’, and *doscientos* [doɾθjentos] ‘two hundred’. Thus, in the case of Spanish, a straightforward reduction in the duration of /s/ produced much lower correct identification rates for /ɾ/ than in Catalan or English. This led the authors to propose that the transitional vocalic element is a possible relevant cue in the perception of coda /ɾ/, a hypothesis tested in the current study.

2.2. Experimental procedure

As stated above, the main objective of this study was to test the perceptual weight of the vocalic element that is present in Spanish /ɾ/+C clusters. In addition, it was hoped that, by doing so, some insight would be gained into related issues, such as the relevance of the duration of /ɾ/ and its exact constriction location for perception. For that purpose, a test was devised that manipulated the duration of an original /s/+C sequence in the hope that a sufficient reduction in length would induce the percept of /ɾ/+C.

[1] See below for full details on this procedure as applied to the current study.

2.2.1. Stimuli

The stimuli consisted of the set of 8 real words given in Table 1. All were disyllabic words of the type $(C_1)VC_2{\cdot}C_3V$, where the raised dot indicates a syllable boundary. The vowel was always /a/; C_1, where present, was /k/; C_2 was either /s/ or /ɾ/; and C_3 was one of the three voiceless stops of Spanish, /p/, /t/ and /k/, or the labial nasal /m/. Stress always fell on the first syllable.

caspa /ˈkaspa/	'dandruff'	carpa /ˈkaɾpa/	'carp'
casta /ˈkasta/	'caste'	carta /ˈkaɾta/	'letter'
casca /ˈkaska/	'hits'	carca /ˈkaɾka/	'conservative'[2]
asma /ˈasma/	'asthma'	arma /ˈaɾma/	'weapon'

Table 1. Stimuli for the perceptual experiment.

One native speaker of Castilian Spanish read each of the above words five times in the carrier sentence *Diga _____ cada vez* 'Say _____ each time'. One token each of *caspa*, *casta*, *casca* and *asma* was chosen as the base signal for subsequent manipulation. Prior to this, average durations for coda /s/ and /ɾ/ were obtained by measuring these segments from the acoustic waveforms of all five tokens. In addition, an average duration was also obtained for the vocalic element that appeared between coda /ɾ/ and the following consonant. Illustrations of these measurements are provided in Figure 2. Table 2 gives average durational values that were obtained for the consonants and the vocalic element.

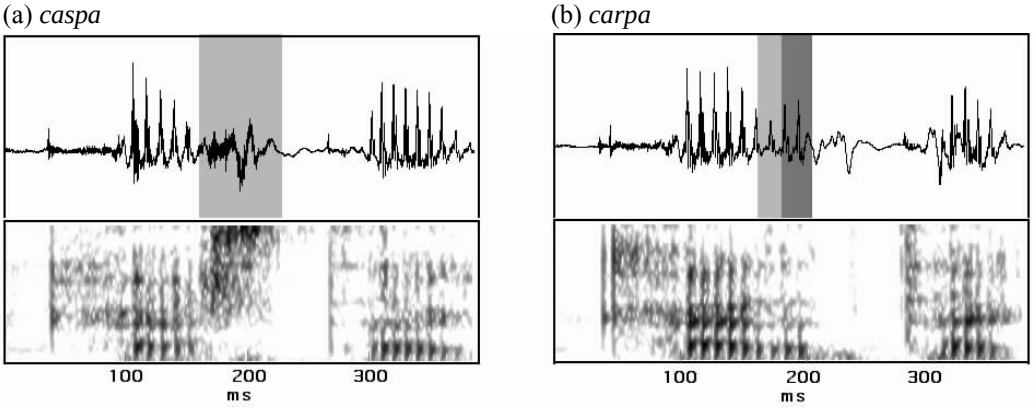

Figure 2. Durational measurements for a token of (a) *caspa* and (b) *carpa*. The shaded areas indicate the portions of the signals corresponding to /s/ in (a) and to /ɾ/ (light shading) and the vocalic element (dark shading) in (b) from which duration values were obtained.

/s/		/ɾ/		Vocalic Element	
Stimuli	Duration	Stimuli	Duration	Stimuli	Duration
caspa	70.4	*carpa*	19.4	*carpa*	25.2
casta	65.5	*carta*	26.0	*carta*	44.7
casca	82.6	*carca*	19.7	*carca*	38.3
asma	72.6	*arma*	48.1	*arma*	51.0

Table 2. Average durations (ms) for the stimuli used in the perceptual study.

Once these values had been obtained, the stimuli to be used in the perception tests were created according to the following procedure. For each pair of words (*caspa-carpa*, *casta-carta*, *casca-carca* and *asma-arma*) three reduction continua were created:

[2] *carca* is a common term for a conservative, old-fashioned person in Spain.

1) continua #1: The token of the s-word that most closely approximated the average duration of /s/ for the given pair was chosen as the base token for subsequent manipulation. The section of the waveform corresponding to the /s/ was identified and its center point located. Starting from the left edge of the /s/ signal, a portion was selected that was equivalent to 50% of the duration of the average /ɾ/ for the word pair, and the same procedure was followed at the right edge. The remaining portion of the /s/ was removed in 25% increments, proceeding from its center point and removing 12.5% both to its left and right. After 100% of the corresponding /s/ signal had been removed, the remaining portion was equal to the average /ɾ/ duration. This ensured that the immediate VC and CV transitions into and out of /s/ were preserved in all tokens. The procedure is illustrated schematically in Figure 3.

2) continua #2: The second set of continua were created following exactly the same steps as for the first with one exception: after the removal of each portion of the /s/ wave, a copy of the vocalic element was added which had been excised from the r-word token that most closely approximated the average duration of the vocalic element for that word pair. The copy of the vocalic element was added at the right edge of the corresponding section.

3) continua #3: The final set of continua were created in the same fashion as the second, the difference being that the copy of the vocalic element that was inserted after the removal of each portion of the /s/ was actually half the duration of the original. This smaller portion of the vocalic element was in turn obtained by removing 50% of its original duration starting from the center point of the signal and removing 25% both to the left and to the right.

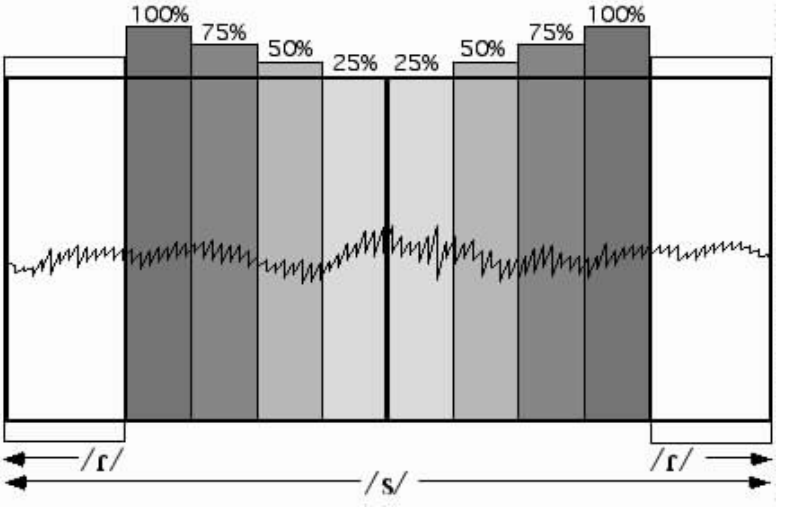

Figure 3. Schematic illustration of how the original set of continua were obtained from the base /s/ signal by identifying a portion of the signal equivalent to the duration of /ɾ/ and then removing portions of the waveform in 25% increments.

	Continua		
	#1: shortening of /s/	**#2: shortening of /s/ plus long vocalic element (vl)**	**#3: shortening of /s/ plus short vocalic element (vs)**
caspa-carpa	*caspa*	*caspa*	*caspa*
	***caspa*-25**	***caspa*-25vl**	***caspa*-25vs**
	***caspa*-50**	***caspa*-50vl**	***caspa*-50vs**
	***caspa*-75**	***caspa*-75vl**	***caspa*-75vs**
	***caspa*-100**	***caspa*-100vl**	***caspa*-100vs**
	carpa	*carpa*	*carpa*

Table 3. List of utterances used in the perceptual experiment for the word pair *caspa-carpa*. Corresponding sets of utterances were used for the remaining 3 word pairs. Utterances resulting from manipulation of the original s-word appear in bold letters.

A total of 12 different manipulated signals were created using this procedure for each word pair. Table 3 lists these signals according to the continuum that they represent for the word pair *caspa-carpa*. Together with the original s-words and r-words, the total number of utterances to be used subsequently in the perceptual experiment was 14 for each word pair. The grand total number of stimuli used in the subsequent perception tests was 56.

2.2.2. Perception test

Each of the 56 stimuli was included twice in a Microsoft Power Point presentation for the perception tests, generating a total of 112 tokens. These tokens were randomized and presented in a XYZ-choice design to the listeners. All of the words appeared in the original carrier sentence *Diga ____ cada vez*. The test was administered to 14 native speakers of peninsular Spanish from the Tarragona area. Accordingly, most of the subjects were Spanish-Catalan bilinguals. However, given the nature of the test and the type of stimuli, this was not considered to be an issue that could negatively affect the outcome of the tests.

The perception tests worked as follows. Subjects were asked to identify each token by indicating what they heard on an answer sheet. For every token the answer sheet provided three options. For example, for the *caspa-carpa* word pair, listeners would hear any of the 18 possible tokens in any of the three continua, and they were asked to identify the token presented as an instance of either *caspa* or *carpa* by circling the appropriate word on the answer sheet. In addition to the words *caspa* and *carpa*, the answer sheet also provided a third column with the option *otra* 'other', which subjects were instructed to choose in the event that the stimulus they heard could not be identified as either of the other options. Subjects were not, however, instructed to provide an alternative for those cases, since it was considered that offering that possibility would distract from the rhythm of the test. In any case, they were told to try to respond as quickly as possible and choose one of the main two options without taking too much time to think about alternatives. The test was self-timed by the listeners, who could choose to listen to any particular stimulus more than once if they felt so. Each subject listened to the stimuli in one single session. The sessions were conducted in a quiet room and the stimuli were presented through headphones.

The results of the test were tabulated and percentages of identification as /ɾ/ for each condition were calculated. For every word pair, identification as /ɾ/ meant that a specific token was identified as an instance of the corresponding r-word. That is, for example, in the case of *caspa-carpa*, the identification as /ɾ/ percentage would be 100 if all instances of the particular *caspa* token (from manipulated stimuli or otherwise) were identified as instances of *carpa*. Given the nature of the stimuli, it was expected that tokens of unmanipulated *caspa*, for example, would yield a percentage of identification as /ɾ/ of 0, while tokens of unmanipulated *carpa* would yield a percentage of identification as /ɾ/ of 100. The manipulated tokens in between were expected to reach higher identification as /ɾ/ rates as the duration of the manipulated signal approached the duration of unmanipulated *carpa*.

The resulting percentages were analyzed using a factorial ANOVA. The dependent variable was percentage of identification as /ɾ/, while the independent variables were the degree of /s/ reduction (-25%, -50%, -75% and -100%) and the duration of the vocalic element (no vocalic element, long vocalic element and short vocalic element). Post-hoc Fisher's PLSD tests were used to check significance between the treatments in the independent variables.

3. Results

As mentioned in the previous section, the results of the perception test were pooled across the 14 subjects and the percentages of /ɾ/-identification were obtained. The raw scores and percentages for all the conditions are listed in Table 4. The percentages shown are presented graphically as line bars in Figure 4 for the word pairs *caspa-carpa, casta-carta, casca-carca* and *asma-arma*. The line bars show, on the horizontal axis, the stimuli presented to the listeners and, on the vertical axis, the percentage of identifications of a specific stimulus as an instance of the r-word in the corresponding word pair.

	caspa-carpa		*casta-carta*		*casca-carca*		*asma-arma*	
	score	%	score	%	score	%	score	%
s	2	7	1	4	1	4	0	0
s–25	3	11	5	18	2	7	0	0
s–50	10	36	1	4	9	32	0	0
s–75	11	39	16	57	15	54	1	4
s–100	17	61	16	57	12	43	15	54
r	28	100	28	100	28	100	28	100
s–25vl	11	40	3	11	6	21	0	0
s–50vl	25	89	15	54	21	75	1	4
s–75vl	27	96	27	96	26	93	10	36
s–100vl	28	100	22	79	28	100	28	100
s–25vs	5	18	4	14	4	14	0	0
s–50vs	14	50	12	43	17	61	0	0
s–75vs	27	96	27	96	27	96	18	64
s–100vs	26	93	28	100	23	82	28	100

Table 4. Perception test results: /ɾ/-identification scores and percentages — all experimental conditions.

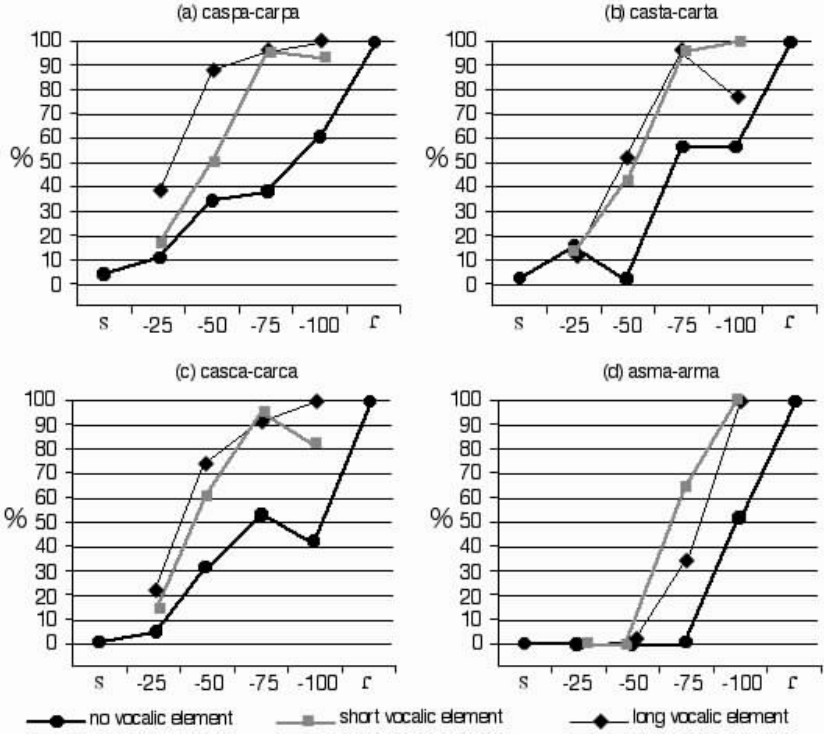

Figure 4. Line graphs for (a) *caspa-carpa*, (b) *casta-carta*, (c) *casca-carca*, (d) *asma-arma*.

The results show an overall pattern that is similar across the different word pairs. It seems clear, from the values in the table and from the observation of the line patterns in the graphs, that it is possible to obtain the percept of /ɾ/ from the simple reduction in duration of /s/ and without an intervening vocalic element. As for the /s/+voiceless stop clusters, some subjects perceived *caspa* as *carpa*, *casta* as *carta*, and *casca* as *carca*, starting with as small as a 25% reduction in the duration of the original /s/. However, only after a reduction of 100% does the perception of the r-words excede the level of chance, except in the case of *casca-carca*, where the 100% reduction results in just 43%

correct identification. In the case of the /s/+nasal cluster, the general trend is the same, except that in this case almost none of the smaller reductions yield any r-percepts at all.

As concerns the presence of the vocalic element, it is clear from inspection of the graphs that it increases the percept of /ɾ/ rather considerably. This is truer for the higher levels of duration reduction, that is, 50%, 75% and 100%, than for the lower level (25%) for which, in most cases, there are no differences at all vis-à-vis the no-vocalic element condition. The results of the ANOVA corroborate these findings. Both main effects, degree of reduction and presence of the vocalic element, were significant ($F(3,36)=28.182$, $p<.01$ and $F(2,36)=12.310$, $p<.01$ respectively). The interaction, however, was not significant. Fisher tests for the vocalic element variable revealed significant differences between the no vocalic element condition and both the long vocalic element and the short vocalic element, both at the .01 alpha level, but not between the long vocalic element and short vocalic element conditions.

4. Discussion

The main hypothesis presented in this paper concerned the role of the so-called svarabhakti vowel in the perception of r+C clusters in Spanish. It was hypothesized that, because this vocalic element is the result of gestural coordination, its presence would not be crucial in the perception of these clusters by native speakers. The results obtained corroborate the hypothesis, although only to some extent. While it is true that it is possible to obtain straightforward r-percepts without the presence of any vocalic element, that is, simply by reducing the duration of an original /s/ to approximate that of an /ɾ/, the addition of a vowel-like portion after the consonant increases the perception of r-tokens significantly. It seems then that the presence of a vocalic element facilitates hearers' perception of r+C clusters even though they are capable of hearing /ɾ/ even when that vocalic element is missing and as long as the duration of the tongue-tip gesture is short enough.

Thus, to the question of whether the vocalic element is essential in the perception of r+C clusters, the answer is negative. This seems in accordance with the articulatory nature of the vocalic element. As proposed by Bradley (2004) the svarabhakti vowel is best understood as the result of gestural coordination. More specifically, as suggested in the introduction, it is the nature of the tongue-tip movement for the /ɾ/ that favors the particular timing between the /ɾ/ and the following gesture, yielding the vocalic element. In other words, the flapping movement of the tongue tip results in a time interval between the contact somewhere in the post-alveolar region (possibly with slight tongue tip retroflexion) and the constriction for the following consonant. It would be assumed then that the tongue-tip gesture associated with Spanish /ɾ/ has strict requirements for length in order to complete the flapping movement. When these requirements are not met, processes like assibilation or consonant gemination arise, as is the case in highland Ecuadorian Spanish or Cuban Spanish, respectively. Figure 5 shows a schematic representation of the formation of the vocalic element from the original tongue-tip traces shown in Figure 1. The circled area in Figure 5 would correspond to the point of contact between the tongue tip and the post-alveolar region during the production of /ɾ/ in the sequence /ɾd̪/ (compare with the much more anterior constriction location for intervocalic /t/, also shown in the figure, which is normally dental in standard Spanish). After contact, the tongue tip slides forward and downward towards the dental point of articulation of /d̪/, a movement indicated by the dashed line. It is this time interval, between the post-alveolar contact for /ɾ/ and the achievement of the dental target for /d̪/, that would be responsible for the formation of the vocalic element. Notice also that, for the single /ɾ/ (dark grey trace), the situation is similar, except that in this case, after hitting the alveolar ridge with a flapping movement, the tongue tip returns to a low position associated with the following vowel. The time interval between the achievement of the post-alveolar target for /ɾ/ and the return to the position for V_2, indicated by the dashed line, would also give rise to a vocalic element. Following this account, a vocalic element would be heard even when /ɾ/ appears in word-final position, as in *cantar*. That is indeed the case, at least judging from an impressionistic point of view. In fact, it is the case that in a highly affected style, this final vocalic element can be heard clearly as a result of an overarticulation of /ɾ/. In the case of r+C clusters, it is hypothesized that the onset of the constriction corresponding to the

subsequent consonant would be timed with the tongue-tip gesture in such a way that there is enough time for the tongue-tip gesture to complete its flapping trajectory.

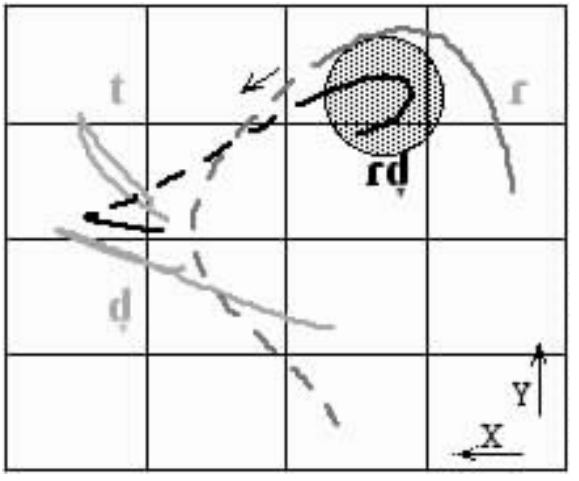

Figure 5. Schematic representation of the generation of the vocalic element in Spanish /rd/ (black dashed trace) and /ɾ/ (grey dashed trace).

Given this epiphenomenal nature of the vocalic element, it is not surprising that Spanish speakers do not find it absolutely necessary for this element to be present in order to perceive /ɾ/. Obviously, its presence increases the identification rate, as we saw earlier, since it represents a much more natural pronunciation and it is probably an important cue in the distinction between /ɾ/ and other tongue tip consonants, such as /t/ or /d/.

The results obtained in this study also reveal the inadequacy of an explanation of the svarabhakti vowel in terms of vowel epenthesis. To begin with, it would seem strange to have an independent process of epenthesis which is not crucial as a cue in the perception of /ɾ/. In other words, if the presence of the epenthetic vowel were somehow rule-governed at the level of phonological planning, then we would expect listeners to rely on its presence more heavily than they seem to. The fact that they do not is indicative that we are not dealing with a process of segmental insertion but rather with a case of variable degrees of articulatory overlap. Moreover, without the articulatory explanation, it would be hard to understand why epenthesis would occur with /ɾ/ but not with other tongue-tip consonants. In addition, the observed dialectal variation mentioned earlier falls out rather naturally if we look at this phenomenon as a consequence of gestural organization, whereas the presence or absence of an epenthetic vowel would not add much to the explanation of these other related phenomena. Still, more detailed work needs to be done in dialects which show blending rather than overlap before the relevance of the articulatory explanation can be stated for these phenomena with certainty. Finally, the similarities between these and other proposed epenthesis explanations, as in the English cases mentioned in the introduction, indicate the soundness of an articulatory explanation.

The finding that there was no significant difference in the degree of correct identification of r-tokens depending on the length of the vocalic element is also indicative that we are dealing with the result of articulatory organization. As shown in Figure 1, the exact trajectory of the tongue tip in the production of /ɾ/ varies depending on factors such as the vocalic context. The figure shows clearly that the flapping movement for /ɾ/ is much wider in the case of a back vowel such as /o/ (right-hand side of the figure) than in the context of a front vowel like /e/ (left-hand side of the figure). A smaller overall movement of /ɾ/ in the context of a front vowel would logically result in a shorter interval between the /ɾ/ and the subsequent consonant, which would yield a shorter vocalic element. Thus, speakers may not rely much on the actual duration of this element, since it varies considerably depending on the context, as indeed shown by the results obtained in this study.

The results obtained in this study also raise some interesting questions concerning the exact articulatory nature of Spanish /ɾ/. In particular, the fact that the presence of the vocalic element significantly increases the percentage of /ɾ/ identification seems to indicate that this is the preferred configuration for native speakers of the language. This means that, whether the exact nature of /ɾ/ is more tap-like or flap-like may not be too important, as long as the duration of the gesture and, perhaps more importantly, the temporal coordination with the surrounding gestures, allow for the formation of this transitional vocalic element. The fact that r-percepts can be obtained from an original /s/ seems to indicate that the exact constriction location of the tongue-tip gesture for /ɾ/ may not be a crucial clue in the perception of this consonant. Instead, we would expect much variability depending on the surrounding vowels as well as the position of the /ɾ/ within the syllable and the word.

5. Conclusion

The present paper has provided evidence that the vocalic element that is commonly heard in r+C consonants in Spanish is best understood, as suggested by Bradley (2004), as the result of the temporal coordination of articulatory gestures. The results of the perceptual tests show that the presence of the vocalic element is not absolutely essential in order to induce the percept of /ɾ/ from a shortened original /s/. This is taken as justification to rule out an explanation in terms of vowel epenthesis. Instead, it opens up interesting possibilities about the exact articulatory nature of /ɾ/ in Spanish and the specifications for constriction location and duration that define this tongue-tip gesture consonant.

Some issues that remain to be investigated in further research would include, among others, the relevance of the laryngeal settings for /ɾ/ and the relationship between consonant duration and voicing. As seen in this study, an originally voiceless consonant like /s/ can be understood as voiced /ɾ/ with sufficient reduction of its duration. It would be interesting to investigate this process further and to see to what extent this relationship holds in other processes of phonological weakening or reduction of consonants, as well as in processes of creation of vocoids from original consonants.

Another issue that remains to be investigated concerns the specific results of the perceptual test used in this study. In particular, it would be interesting to determine whether listeners can tell the difference between an original /ɾ/ token and the /s/ tokens that are understood as instances of /ɾ/. A discrimination test would be necessary in order to address this issue, but, at least judging from the current results and from the very few cases in which the majority of the listeners that participated in the experiment opted for the *otra* option, it would seem that they did not perceive much of a difference between the two sets of stimuli. Nonetheless, this particular aspect remains to be tested experimentally.

One final issue that deserves some comment, as far as the results of the perceptual tests are concerned, is the distinct behavior observed in the sequences with a nasal consonant, i.e., the *asma-arma* word pair. The fact that for this particular condition a much more radical reduction in the duration of /s/ was necessary before any r-percepts were obtained at all might be related to the voicing issue. The nasal /m/, as opposed to the other consonants used in the study, is voiced, which, under normal circumstances in Spanish, would trigger voicing assimilation of the preceding /s/ in the sequence *asma*. Even so, one would expect the fact that the original /s/ is somewhat voiced to favor the percept of /ɾ/ even with less reduction in the duration. The fact that this is not so points at the need to investigate other aspects of /m/, such as the necessary extra complexity added by the velum-lowering gesture, as a possible source of this difference with the voiceless stops.

Acknowledgments

I would like to thank John Kingston for his useful comments on the perceptual test during the LASP3 conference and María Riera for her comments on an earlier version of this paper.

This study has been partly supported by project 2005SGR864 (Grup de Recerca Consolidat en Fonètica Experimental; Generalitat de Catalunya) and by grant HUM2005-02746 from the Ministry of Science and Technology, Spain.

References

Argüello, Fanny (1978). *The Zeísta Dialect of Spanish Spoken in Ecuador: A Phonetic and Phonological Study*. PhD dissertation, The Pennsylvania State University.

Bradley, Travis (2004). Gestural timing and rhotic variation. In Face, T. (Ed.), *Laboratory Approaches to Spanish Phonology* (pp. 195-220). Berlin: Mouton de Gruyter.

Browman, Catherine, & Goldstein, Louis (1986). Towards an anticulatory phonology. *Phonology Yearbook* (3), 219-252.

---------- (1989). Articulatory gestures as phonological units. *Phonology* (6), 201-251.

---------- (1992a). Articulatory Phonology: An overview. *Phonetica* (49), 155-180.

---------- (1992b). Targetless schwa: An articulatory analysis. In Docherty, G., & Ladd, R. (Eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody* (pp. 26-56). Cambridge: Cambridge University Press.

Colantoni, Laura, & Steele, Jeffrey (2005). Phonetically-driven epenthesis asymmetries in French and Spanish obstruent-liquid clusters. In Gess, R., & Rubin, E. (Eds.), *Theoretical and Experimental Approaches to Romance Linguistics* (pp. 77-96). Amsterdam: John Benjamins.

Fowler, Carol (1990). Some regularities in speech are not consequences of formal rules: Comments on Keating's paper. In Kingston, J., & Beckman, M. (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 476-489). Cambridge: Cambridge University Press.

---------- (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America* (99), 1730-1741.

Gili Gaya, Samuel (1921). La r simple en la pronunciación española. *Revista de Filología Española* (8), 271-280.

Granberry, Julian (1999). *Essential Swedish Grammar*. New York: Dover Publications.

Guitart, Jorge (1976). *Markedness and a Cuban Dialect of Spanish*. Washington, DC: Georgetown University Press.

Hualde, José Ignacio (2005). *The Sounds of Spanish*. Cambridge: Cambridge University Press.

Kingston, John, & Diehl, Randy (1994). Phonetic knowledge. *Language* (70), 419-454.

Liberman, Alvin, & Mattingly, Ignatius (1985). The motor theory of speech perception revised. *Cognition* (21), 1-36.

Lindblom, Björn (1996). Role of articulation in speech perception: clues from production. *Journal of the Acoustical Society of America* (99), 1683-1692.

Malmberg, Bertil (1965). *Estudios de Fonética Hispánica*. Madrid: Consejo Superior de Investigaciones Científicas.

Martínez Celdrán, Eugenio (1978) *Fonética*. Barcelona: Teide.

Navarro Tomás, Tomás (1970). *Manual de Pronunciación Española*. Madrid: Consejo Superior de Investigaciones Científicas.

Olive, Joseph, Greenwood, Alice, & Coleman, John (1993). *Acoustics of American English Speech. A Dynamic Approach*. New York: Springer.

Quilis, Antonio (1970). El elemento esvarabático en los grupos [PR, BR, TR...]. In Straka, G. (Ed.) *Phonétique et Linguistique Romanes, I: Mélanges Offerts à M. Georges Straka* (pp. 99-104). Strasbourg: Société de Linguistique Romane.

Ramírez, Carlos Julio (2006). Acoustic and perceptual characterization of the epenthetic vowel between clusters formed by consonant+liquid in Spanish. In Díaz-Campos, M. (Ed.), *Selected Proceedings of the Second Conference on Laboratory Approaches to Spanish Phonology* (pp. 48-61). Somerville, MA: Cascadilla Proceedings Project.

Ravelo, Pedro (1999). El llamado "elemento esvarabático": ¿sólo un curioso segmento vocálico?. *Actas del I Congreso de Fonética Experimental*. Barcelona, Spain: Publicacions de la Universitat de Barcelona.

Recasens, Daniel (submitted). Phonetic typology and positional allophones for alveolar rhotics in Catalan. *Phonetica*.

Riera, María, & Romero, Joaquín (2006). V+/l/ and V+/r/ sequences in American English: a preliminary acoustic study. In *Proceedings of the 29th AEDEAN International Conference* (pp. 529-536). Jaén, Spain: Universidad de Jaén, Servicio de Publicaciones.

Romero, Joaquín (1995). *Gestural Organization in Spanish: An Experimental Study of Spirantization and Aspiration*. PhD dissertation. University of Connecticut.

---------- (1996). Articulatory blending of lingual gestures. *Journal of Phonetics* (24), 99-111.

---------- (2001). Temporal reduction effects in diachronic change: rhotacism. Paper presented at Phonetics-Phonology Interface conference. ZAS, Berlin, Germany. October.

Romero, Joaquín, & Martín, Sidney (2003). Articulatory weakening as basis of historical rhotacim. In Solé, M. J., Recasens, D., & Romero, J. (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2825-2828), Barcelona, Spain. Melbourne: Causal Productions.

Schmeiser, Benjamin (2006). Durational variability of svarabhakti vowels in Spanish complex onsets: a multi-country perspective. Paper presented at the 3rd Laboratory Approaches to Spanish Phonetics and Phonology conference. University of Toronto, Toronto, Canada. September.

Wells, John (2000). *Longman Pronunciation Dictionary* (New Edition). Harlow, England: Longman-Pearson Education.

Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology

edited by Laura Colantoni and Jeffrey Steele

Cascadilla Proceedings Project Somerville, MA 2008

Copyright information

ISBN 1-57473-424-9 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Romero, Joaquín. 2008. Gestural Timing in the Perception of Spanish r+C Clusters. In *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, ed. Laura Colantoni and Jeffrey Steele, 59-71. Somerville, MA: Cascadilla Proceedings Project.

or:

Romero, Joaquín. 2008. Gestural Timing in the Perception of Spanish r+C Clusters. In *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, ed. Laura Colantoni and Jeffrey Steele, 59-71. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1714.