PERCEPTUAL STRATEGIES DUE TO TRILL-VOWEL COARTICULATORY EFFECTS IN SPANISH

Julie A. Lewis
University of California, Berkeley, USA

ABSTRACT

In Spanish, the alveolar trill affects the articulation and resulting acoustic realization of certain preceding vowels. For example, it creates a lower allophone of the close-mid front vowel /e/. To further elucidate the role of coarticulatory effects in perception, a perceptual experiment was conducted to test whether listeners can use such vocalic cues to figure out the identity of the following consonant. After hearing the first half of the words *pero* “but” and *perro* “dog” (/pe-/), listeners were indeed able to identify the second half of the word at a level significantly above chance, given a forced choice between *-ro* (/-ɾo/) and *-rro* (/ro/). It appears they can use allophonic differences in the vowel due to coarticulation with a trill as cues for the trill. This finding has implications for second language acquisition, speech pathology, and speech technology.

1. INTRODUCTION

1.1. Spanish Coarticulation

Cross-linguistically, r sounds affect the vowels that precede them, for example, lengthening or lowering them [1], and Spanish is no exception. Quilis's data show that Spanish /e/ before the alveolar trill is lower and more back (having a higher F1 and a lower F2) than it is before other sounds (p. 157, female speaker) [2]:

	F1	F2
bebe	324(Hz.)	2.146
cerveza	283	2.025
perro	**405**	**1.822**
negras	283	2.349
trece	283	2.106

Table 1. Formants of /e/ before different consonants (from Quilis).

Descriptions of the articulation of alveolar trills point to a likely reason for this acoustic effect. Trills ([r]) involve the tongue tip being raised toward the alveolar ridge in a manner such that when air is passed through the constriction, the tongue tip vibrates, creating a succession of brief alveolar closures; the tongue body needs to be lowered in order to raise the tongue tip to make the trill. The tap ([ɾ]) also involves an alveolar closure, but it is articulated via one ballistic movement of the tongue tip and does not require tongue body lowering. Recasens found less palatal contact in [iri] and [uru] than in [iɾi] and [uɾu] in his electropalatographic data for Catalan and suggests that trills work best with the tongue's predorsum and mediodorsum lowered, the sides fixed to help the tip and blade make the required closure (consistent with McGowan's ideas [3]), and the postdorsum retracted [4]. This lowering of the tongue body may well create lowered vowels (acoustically). Furthermore, Recasens found that, in general, there is less of a vowel to consonant effect between [r] and surrounding vowels than there is for [ɾ], and that [r] affects [i] more than [i] affects it; i.e., in the battle between trills and surrounding vowels, trills win. Thus, for articulatory reasons, trills have a strong effect on the vowels that precede them, in particular, lowering the vowels that generally involve the tongue body being raised. (Since the tongue body is already lowered, lowered allophones are not expected -or found- for low vowels before trills.)

1.2. Coarticulation and Perception

Speech segments are not perceived discretely, or independently of the segments that surround them. Coarticulatory effects such as those described above have perceptual ramifications. A large body of literature reports that listeners are aware on some level of coarticulatory effects and use them perceptually. For example, Fowler’s experiments on VCV coarticulatory effects in English show that listeners hear acoustically identical vowels as different when the vowels that surround them are different and hear acoustically different vowels (vowels with different coarticulatory influences) as the same when each vowel is in its appropriate vocalic context [5]. Listeners factor in the coarticulatory effects that they know occur in speech, using them as cues to figure out the identity of affected segments. Widdison invokes these listener strategies to explain certain sibilant patterns in Spanish: “The nature of sounds and how they interact in speech with properties of other segments must constitute an integral part of what speakers know about language” [6]. Listeners use this knowledge to factor coarticulatory effects out, but sometimes fail to do this normalization and mini-sound changes occur (Ohala, [7]).

However, listeners could use this knowledge of how sounds interact not just to factor out effects and figure out the intended identity of the affected phoneme, but also as cues to the identity of the affecting/altering segment, in essence working backwards. For instance, when in a VC or CV sequence the consonant colors the vowel, listeners know this and so factor out the effects to correctly perceive the vowel, but they could also use the effects on the vowel as cues to the identity of the consonant, tracing the perturbance of the vowel back to its source.

Just such a process has been shown to occur in the perception of nasal consonants in Italian. Maturi found that the nasality of a coarticulated vowel following a nasal consonant is a bigger cue to the nasality of the consonant than the nature of the consonant itself [8]. Subjects did not judge tokens of *strada* with a final nasalized vowel (spliced from *strana*) to be the same as *strada*, nor did they judge *strana* with a final oral vowel (taken from *strada*) to be different from *strada*, at least within a forced-choice setting.

Another instance of vowels affected by a surrounding consonant providing cues to the identity of that consonant can be seen in Whalen’s experiments showing that the formant transitions of the vowel in fricative-vowel sequences in English provide cues to the place of articulation of the fricative [9]. For example, for the same fricative segment (with a particular spectrum of noise), there were many fewer “sh” (vs. “s”) responses when the transition appropriate to [s] was present in the vowel than when the transition was appropriate to [ʃ]. Similarly, Fischer and Ohde found that the duration and the offset frequency of the F1 transition of the preceding vowel affect the perception of the categories /k/ vs. /g/, velar stops, word finally [10]. For instance, a longer duration of the vowel led to higher rate of perception of /g/. Warren and Marslen-Wilson explore how anticipatory coarticulatory effects are used in spoken word recognition, suggesting that such effects are used on-line for the fastest possible identification of words [11].

1.3. The Present Experiment

As the above studies show, if consistent coarticulatory effects occur in a language, listeners may well be using them perceptually. If Spanish trills' lowering effects on vowels have created allophonic differences in the (mid) vowels that precede them, listeners could be using these vocalic differences as cues to the identity of the consonant. This experiment tests whether or not speakers can indeed use these coarticulatory effects in the close-mid-front vowel /e/ in Spanish as cues to the identity of the upcoming consonant, just as Italian listeners use nasal-vowel coarticulatory effects in the vowel as cues to the identity of the consonant, as shown in Maturi's experiment cited above. To test this idea, I will see if listeners can predict the following consonant after hearing various allophones of /e/ from real speech.

2. METHODS

A perceptual experiment was conducted, playing the first half of tokens of the Spanish words *perro* “dog” and *pero* “but” (i.e., [pɛ-] and [pe-]) to see if native speaker listeners could determine the second half when given a forced choice between (orthographic) “-rro” or “-ro” ([-ro] and [-ɾo], respectively). *Carro* [karo] “car” and *caro* [kaɾo] “expensive” were used as a control. The hypothesis was that listeners would be better at determining the following consonant for the words with the mid vowels than for the words with the low vowels, where there is not a consistent allophonic difference before the trill and the tap. Listeners should be able to predict the consonant after the affected vowel (the lowered allophone before the trill)–and after the unaffected one (the allophone before the tap)–with some success, as there are usable cues to the consonant’s identity in the vowel. The expectation for the low vowels was that listeners would perform at the chance level, as there are not such cues in them.

2.1. Stimuli Preparation

A female native speaker of Iberian Spanish was recorded six times (digitized at a sampling rate of 16,000 samples per second) speaking the words *pero, perro, caro* and *carro* in the frame sentence *Di ___dos veces* (“Say ___ two times”). The acoustic measurements (in Hz.) of the vowels in the first syllables follow. Formants were obtained via LPC analysis (filter order 18).

	F1	F2
1pero	452	2370
2pero	495	2262
3pero	452	2305
4pero	463	2348
5pero	463	2348
6pero	495	2273
Average	**470**	**2318**
1perro	603	2047
2perro	603	2133
3perro	614	2090
4perro	592	2197
5perro	571	2122
6perro	538	2187
Average	**587**	**2129**

Table 2. Formants of /e/ before the tap and the trill.

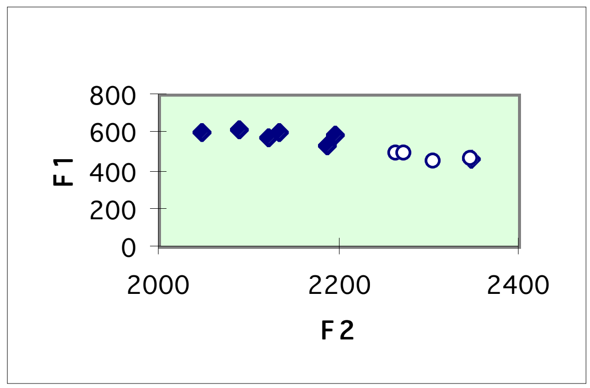

Figure 1. Plot of F1 and F2 for /e/ before taps and trills. Open circles are /e/ before the tap, and solid diamonds are /e/ before the trill.

The mid vowel /e/ is lower and more back in all six pairs (replicating Quilis’ data). The allophones do not overlap in the vowel space (see Figure 1). No such consistent difference was found for the (already) low vowel /a/:

	F1	F2
1caro	926	1853
2caro	809	1853
3caro	937	1781
4caro	840	1840
5caro	818	1874
6caro	840	1853
Average	**862**	**1842**
1carro	905	1823
2carro	885	1820
3carro	818	1831
4carro	980	1865
5carro	883	1863
6carro	808	1820
Average	**880**	**1837**

Table 3. Formants of /a/ before the tap and the trill.

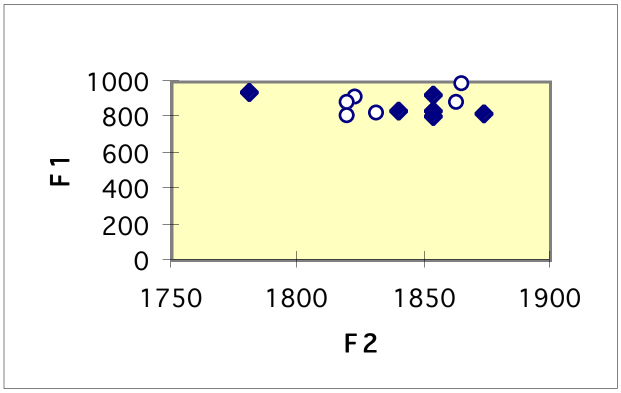

Figure 2. Plot of F1 and F2 for /a/ before taps and trills. Open circles are /a/ before the trill, and solid diamonds are /a/ before the tap.

Vowel length was not significantly different before the tap and trill for *pero/perro* or *caro/carro*. Experimental tokens were created by splicing the target words between the first vowel and the tap or trill, cutting at the decrease in amplitude at the beginning of the consonant in the waveform (leaving in the vowel transitions). The spliced portion from each of the six utterances was presented twice, creating 12 tokens of [pe-], [pɛ-], [ka-] before the tap, and [ka-] before the trill.

2.2. Stimuli Presentation

Fifteen volunteer native Spanish speaker subjects (from various dialects) participated in the two blocks of the experiment (the /e/ and /a/ presentations). Tokens were presented in random order within each block over headphones in a quiet environment. Subjects were instructed to listen to the first half of the words (*pero* and *perro* or *caro* and *carro*) and then to click on one of two boxes on the computer screen, labeled *-ro* and *-rro*. They heard each token only once, but were allowed to change their answers before moving on to the next token.

3. RESULTS

The subjects' responses uphold the hypothesis. For the thirteen subjects whose responses are included in the analysis[1], they correctly identified the upcoming consonant for the mid-vowel tokens 68.91% of the time, a percentage that is much higher than what would be expected if they were guessing, having no cues to the consonant's identity in the vowel. A test of significance (a t-test) showed that this result is not due to chance ($p < 0.01$). As predicted, subjects performed at a level near chance for the low vowel tokens, correctly identifying the upcoming consonant 52.56% of the time. Furthermore, the /e/ tokens were always presented before the /a/ ones, so if there was a diminished performance level as the subjects became accustomed to the experimental format, it would have affected the /e/ responses and not the /a/ ones.

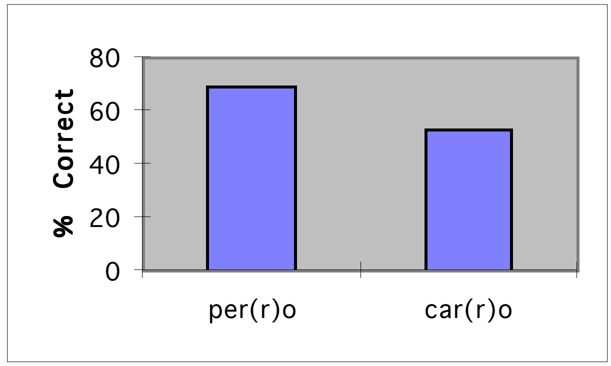

Figure 3. Percent correct for /e/ and /a/ tokens. Line represents 57%, the level up to which results could be due to chance (0.01 significance level).

4. DISCUSSION

The results suggest that Spanish listeners can indeed use the coarticulatory effects that trills create in the close-mid front vowels as cues to the trill's identity. However, the most salient cues to the identity of a consonant are usually in the consonant itself. The primary cue to the identity of a trill vs. a tap in Spanish is probably the presence of more than one stop gap (period of silence) in the acoustic signal. The notion of trading relations may shed light on how different cues can work together in perception. Repp discusses how when several signal properties are perceived together as a phonetic distinction, they are all cues, but they may differ in their relative importance [12]. In fact, the relative importance of the two cues may change in a manner such that they trade off against each other: "If one cue in such an ensemble is changed to favor category ***B***, another cue can be modified to favor category ***A***, so that the phonetic percept remains unchanged. This is called a trading relation." Similarly, as Hodgson and Miller discuss, if a property contributing to the identity of a phonetic contrast weakens (rather than changing), a trading relation can result between it and another (secondary) cue: "...if one of the properties is weakened, so as to render the percept ambiguous, compensation for such a weakening (within limits) can be effected, and the original percept recovered, by the provision of a stronger contribution from the other property. Compensation is contingent upon the listener being sensitive to both acoustic properties" [13][2]. Hodgson and Miller suggest that there are limitations as to how much one of the cues can be weakened, but in this experiment, the cues from the trill were completely obscured, since the listener did not hear any part of the consonant following the vowel.

Thus, in creating the percept of (the category of) the alveolar trill in Spanish, it appears that, at least when it occurs after the close-mid front vowel /e/, the cues from the trill and the effects of the trill on the preceding vowel (in effect, a lowered allophone) are in a trading relationship. The results of this experiment show that as the role of the trill diminishes, here, down to nothing, the vocalic cues can compensate and the trill can still be perceived, at least to a certain extent and within a forced-choice setting. Listeners are indeed sensitive to the allophonic variations within /e/ and can use them as (partial) cues to what consonant is coming.

Due to this trading relation, as the vowel plays a larger role in cueing the identity of the trill, the trill itself could shoulder less of the burden in establishing its own identity. There is some evidence for this outside the laboratory. There are native speakers of Spanish who cannot make an alveolar trill. Since these speakers cannot produce the cues for the trill that come from the trill itself, they could exploit the vocalic cues to create the distinction in words such as *pero* and *perro*.

The vowel quality in these words was analyzed (in Hz.) for two non-trilling native Spanish speakers:

	pero		*perro*	
	F1	F2	F1	F2
Speaker 1	515	2198	446	2198
(adult female)	480	2301	412	2129
			412	2060
			446	2060
Average F2		**2249.5**	**2111.75**	
Speaker 2	584	2644	515	2369
(12-year-old	584	2679	515	2507
female)	549	2610	480	2438
	652	2610	547	2499
	652	2747	515	2438
	547	2499	549	2541
			480	2507
Average F2		**2631.5**	**2471.3**	

Table 4. Formants of /e/ in *pero* and *perro* for non-trilling native Spanish speakers.

While F1 is not consistently higher in *perro* (before the trill) for either subject, F2 is lower, as expected. The average of F2 is 2111.75 Hz. before the trill versus 2249.5 before the tap for speaker one, and 2471.3 before the trill versus 2631.5 before the tap for speaker two. These speakers appear to be recreating the backness (but not the lowness) of the /e/ allophones before the trill. While the data are not entirely conclusive it does seem to indicate that as native speakers (and hearers) of Spanish, these individuals are aware on some level of the allophonic difference that the trill creates and are making use of it in some way as they distinguish *perro* from *pero* (in production) without making a trill.

5. CONCLUSION

The vowel quality in the close-mid front vowel /e/, lowered due to coarticulation with a following trill, does some of the work in establishing the identity of the following trill in Spanish. Listeners are indeed in some way aware of these coarticulatory effects in the vowel and use them as cues to the identity of the following consonant. Perception of the trill begins before the trill's first closure. This knowledge of how trills can affect preceding vowels and how the effects are salient perceptually has applications in both speech pathology and second language learning. When someone is unable to make an alveolar trill, a relatively difficult consonant to produce, he or she can provide some of the trill cues by using the characteristic (lower) allophone of the vowel. Furthermore, speech technology could incorporate this allophony in vowels in appropriate contexts to create and recognize (more) natural sound sequences.

ACKNOWLEDGMENTS

I would like to thank María-Josep Solé for all her help in the data collection and analysis for this study.

NOTES

1. Two subjects' reponses were not included in the statistical analysis because they did not respond to all of the tokens. However, their response patterns were similar to those of the other subjects.
2. Repp and Hodgson and Miller disagree about whether or not trading relations occur within phonetic categories, but they agree that they happen between categories, to create categorical distinctions.

REFERENCES

[1] Lindau, M. 1985. The story of /r/. In V. A. Fromkin (ed.), *Phonetic Linguistics*. Orlando, FL: Academic Press. 157-168.

[2] Quilis, Antonio. 1981. *Fonética acústica de la lengua española*. Madrid: Editorial Gredos, S.A.

[3] McGowan, R.S. 1992. Tongue-tip trills and vocal-tract wall compliance. *Journal of the Acoustical Society of America,* 91, 2903-2910.

[4] Recasens, Daniel. 1991. On the production characteristics of apico-alveolar taps and trills. *Journal of Phonetics,* 19, 267-280.

[5] Fowler, Carol A. 1981. Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research,* 46, 127-139.

[6] Widdison, Kirk. 1997. Phonetic explanations for sibilant patterns in Spanish. *Lingua*, 102, 253-264.

[7] Ohala, J. J. 1993a. The phonetics of sound change. In Charles Jones (ed.), *Historical Linguistics: Problems and Perspectives*. London: Longman. 237-278.

[8] Maturi, Pietro. 1991. The perception of consonantal nasality in Italian: Conditioning factors. In *Proceedings of the XIIth International Congress of Phonetic Sciences*. Provence: Universit´ de Provence Service des Publications.

[9] Whalen, D.H. 1981. Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary. *Journal of the Acoustical Society of America,* 69(1), 275-282.

[10] Fischer, Rebecca M. and Ralph N. Ohde. 1990. Spectral and duration properties of front vowels as cues to final stop-consonant voicing. *Journal of the Acoustical Society of America,* 88(3), 1250-1259.

[11] Warren, Paul and William Marslen-Wilson. 1987. Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics,* 41(3), 262-275.

[12] Repp, Bruno H. 1983. Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. *Speech Communication*, 2, 341-361.

[13] Hodgson, Philip and Joanne L. Miller. 1996. Internal structure of phonetic categories: Evidence for within-category trading relations. *Journal of the Acoustical Society of America,* 100(1), 565-576.