Gesture overlap in Portuguese lexical and post-lexical consonant clusters

Conceicao Cunha[1,2], Jonathan Harrington[1]

[1]IPS – Institute of Phonetics and Speech Processing
Schellingstr. 3/ II, 80799 Munich, Germany

[2]LIPP – International Doctoral Program in Linguistics
Schellingstr. 10, 80799 Munich, Germany

cunha|jmh@phonetik.uni-muenchen.de

Abstract. *This study focuses on the extent to which differences in the timing and overlap patterns are to be observed between lexical ([krer]) and post-lexical ([k(ɨ)rer]) clusters in European Portuguese. For the timing measurements the following three variables were analyzed: gesture overlap, plateau lag and target latency. The results showed consistently longer lags, and therefore a wider phasing in post-lexical clusters, independently of position of the cluster in the word and identity of C1. Overall, lexical and post-lexical clusters exhibited small but consistent differences in timing, suggesting that the contrast between them is still maintained.*

1. Introduction

European Portuguese high vowels /ɨ,u/ in unstressed position are mostly deleted in connected speech. This vowel deletion is more frequent in final and medial than in initial position and occurs mainly in CV(C) syllables (Mateus & Martins, 1982; Martins et al, 1995; Silva, 2007; Cunha, *in press*). A consequence of this high-vowel deletion is the formation of post-lexical (PL) clusters that are similar on the surface to lexical (L) clusters, resulting in near homophones (e.g. PL /k(ɨ)rer/, "to want" and L /krer/, "to believe"). It is the (dis)similarity of these PL and L consonant clusters that is the main concern of this study.

2. Perception

2.1. Method

As part of a physiological study (cf. Section 3), audio data were recorded with a multichannel DAT device from two native speakers of European Portuguese. The subjects were 25 and 28 years of age and both from Porto.

The relevant stimuli for this study consisted of Portuguese words containing /pr/, /pɨr/, /kr/, /kɨr/ in initial and in medial position (Table 1), read 3-4 times in random order from a computer screen.

	Position	L		PL	
/pr/	**initial**	*pr*ece	‘prayer’	*per*ece	‘spoils’
		*pr*ecede	‘precedes’	*per*ecer	‘to spoil’
	medial	es*pr*eme	‘squeezes’	es*per*ar	‘to wait’
		es*pr*emer	‘to squeeze’	es*per*ançado	‘hopeful’
/kr/	**initial**	*cr*er	‘to believe’	*quer*er	‘to want’
		*cr*idinho	‘creditor’	*quer*idinho	‘dear’
	medial	es*cr*ever	‘to write’	des*quer*er	‘do not believe’
		es*cr*evinhar	‘to scribble’	des*quer*ido	‘unpopular’

Table 1. Target words. The relevant sequences are in *italics*.

The speakers produced the target words embedded in two carrier sentences, one with a prosodically accented context and one with a prosodically deaccented context: *O Pedro leu ___mal* ‘Peter read ___wrong’ and *O Tiago leu ___mal? –Não, o PEDRO leu ___ mal* ‘Did Tiago read ___wrong? - No, PETER read ___wrong’, respectively.

The target clusters of one of the speakers chosen randomly were excised between their acoustic onset and offset and presented in randomized order in an online forced choice experiment. The participants listened to each stimulus separately and carried out an identification task in which they responded to each /p/-initial token with one of <pr>, <pre>, <per>, <pur> and to each /k/-initial token with one of <kr>, <kre>, <ker>, <kur>. Since the main research question in this study was whether subjects heard a cluster or a CVC-sequence, the first two and the last two choices were pooled for the analysis.

Twenty-one native speakers of European Portuguese - EP (from Porto, and aged between 24-36 years) and sixteen native speakers of Brazilian Portuguese - BP (from Campinas, and aged between 22-35 years) participated in the perception experiment. None of the subjects reported any hearing or reading problems. The responses were analyzed with a generalized linear mixed model (GLMM) in the R statistical program with the listener as a random factor and listener variety as fixed factor.

2.2 Results

The results, plotted in Figure 1, showed that the stimuli were overwhelmingly (>80%) perceived as a vowel-less cluster (/Cr/) irrespective of the intended production (L or PL) and the native language of the participants (EP or BP). EP participants perceived a vowel in the PL-clusters in 21% of the cases, and BP participants in 15% of the cases. The statistical analysis (GLMM) showed however that the pattern of responses differed between L and PL stimuli ($z = 5.1$, $p < 0.001$). Further analyses by variety showed that only EP ($z = 6.8$, $p < 0.001$) but not BP listeners distinguished perceptually between L and PL clusters.

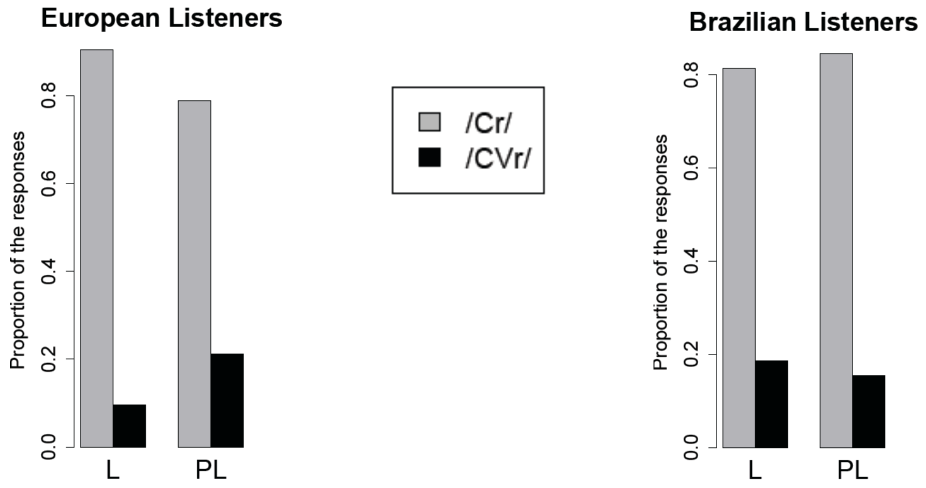

Figure 1. Perceptual responses, presented by cluster type (L vs. PL) for European and Brazilian listeners.

2.3 Discussion

Overall, this perception experiment showed that Brazilian listeners were not able to distinguish between L and PL clusters when produced by European speakers, whereas EP listeners performed this differentiation slightly better. This suggested that there was some difference in the signal that at least EP listeners could use to distinguish between L and PL clusters. In order to determine what information was available to the EP listeners, we collected physiological data, with the aim of testing whether there were any differences in the coordination of the consonants gestures in L and PL sequences. The results of this experiment are presented next.

3. Physiological study

Previous studies have shown that the contrast between pairs differing in the presence or absence of a pretonic unstressed vowel were not completely neutralized (e.g., *beret*/*bray*; *police*/*please*) and that the difference between the words was maintained phonetically because of the longer lag between the two consonants in the words with lexically unstressed vowels (Browman & Goldstein, 1990; 1992; Geng et al., 2010). In light of the perception results, we therefore expected that slightly different timing patterns may also characterize lexical and post lexical clusters in EP.

We further aimed to assess whether additional conditions affected cluster timing and the probability of vowel reduction. Specifically, we tested whether the place of articulation of the first consonant in the cluster (C_1) affected the timing of L and PL clusters. We predicted a greater overlap in the bilabial variants than in the velar ones because of the action of independent articulators by the production of /p/.

Finally, we investigated the stability of L and PL timing as a function of position of the target cluster in the word, comparing clusters initially and medially. Since onset clusters can occur in utterance initial position, where there would be no acoustic cues from a preceding vowel as to the identity of C1, word-initial clusters may be timed with a wider lag in order to preserve the inherent information of C_1 (Chitoran et al, 2002). Consequently, we expected more overlap medially than initially.

3.1 Method

Movement data were recorded using a 5D electromagnetic articulograph at the IPS, Munich (AG500, Carstens Medizinelektronik, cf. Hoole & Zierdt, 2010) from two first language speakers of European Portuguese. The recordings were carried out with the sensors fixed mid-sagittally on the upper and lower lips, on the jaw, and on the tongue tip, tongue body, and tongue back/dorsum. Reference sensors were also attached on the maxilla, nose and on the left and right mastoid bones.

The data were automatically segmented and labeled on the basis of the acoustic signal using the Munich Automatic Segmentation System (MAUS, Schiel, 1999); the labeling of the physiological data and subsequent analyses were made using the EMU System (Harrington, 2010). For the physiological annotation of /p/, the lip aperture (la) was calculated as the Euclidean distance between upper and lower lip sensors. The other two consonants were defined based on the vertical movement of tongue tip (tt) for /r/ and tongue back (tb) for /k/. Figure 2 shows an example of /k/ with the movement of tongue back in mm (above) and the vertical velocity of this articulator in mm/s (below).

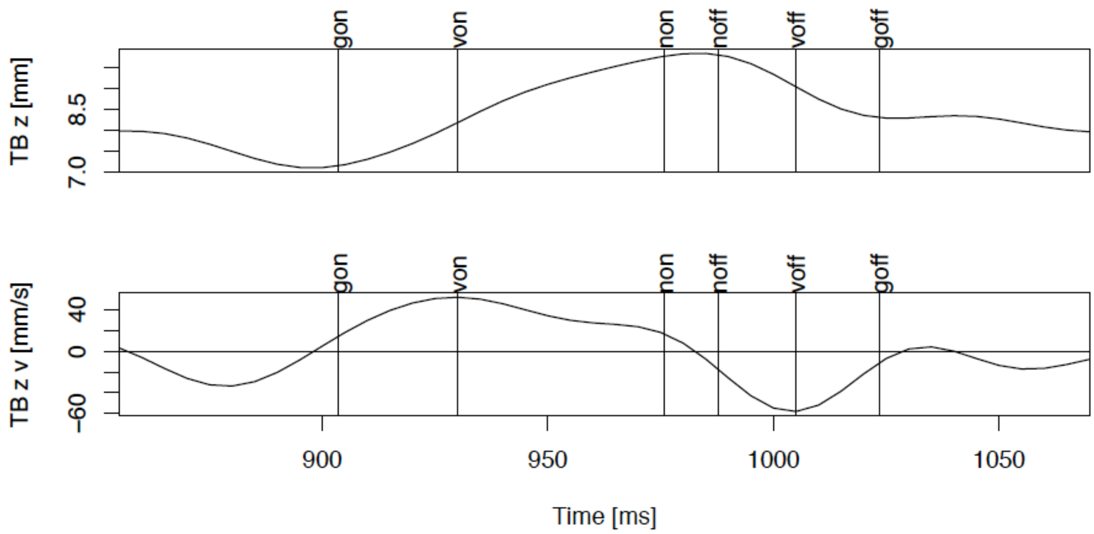

Figure 2. Vertical tongue back position (above) and velocity (below) of /k/ in the token *crer* ('believe', speaker 2, Token 138): gesture onset (GON), maximum onset velocity (VON), target (NON), maximum constriction (MON), release (NOFF), Maximum offset velocity (VOFF), Gesture offset (GOFF).

Maximum onset velocity (VON), maximum constriction (MON) and maximum offset velocity (VOFF) were determined based on changes in the velocity profile of lip aperture/vertical tongue movement. VON and VOFF correspond to the kinematic event with the maximal velocity at the beginning and end of the gestures respectively and MON to the event with the minimal velocity. Gesture onset (GON), Gesture offset (GOFF) as well as the beginning and end of the constriction plateau, coinciding with the achievement of the target and release of the consonant (NON and NOFF) were interpolated values and represent the 20% threshold of the difference between two adjacent extrema in the velocity signal (Bombien, 2011:62). NON was the point in time at which velocity fell below 20% of the range between preceding velocity peak VON and minimum velocity MON; NOFF was the point in time at which velocity exceeded 20% of the range between MON and VOFF (Bombien, 2011; Marin & Pouplier 2010).

The variables used to determine relative timing of the consonants in the clusters were as follows:

a) GESTURE OVERLAP (GOVER) defined as the interval between the end of the first consonant (C_1) gesture and the beginning of the second consonant (C_2) gesture ($C_{1.goff}$ - $C_{2.gon}$, as in Kühnert et al, 2006). A positive value means an overlapping interval of the consonant gestures and a negative value a lag between them.

b) PLATEAU LAG (PLAG) defined as the interval or the lag between release of C1 and achievement of target of C2 ($C_{1.noff}$ - $C_{2.non}$, as in Chitoran et al, 2002; Kühnert et al, 2006, Bombien, 2011:65). As in the gesture overlap above, a positive value means an overlap, in this case of the constrictions plateaus, and a negative value a lag between both plateaus.

c) TARGET LATENCY (TLAT) defined as the time interval between achievement of target of C_1 (NON) and achievement of target of C_2 ($C_{2.non}$-$C_{1.non}$, as in Bombien, 2011:65). This variable defines how close in time the two consonants achieve their targets relative to each other: a positive value means an overlap between the two targets and a negative value a lag between them.

Analyses of variance (ANOVA) of each variable were calculated separately for each speaker in R. The independent variables were lexicality, place of articulation and position.

3.2 Results

3.2.1 Lexicality

PL clusters showed a smaller gesture overlap than L clusters for both speakers. Both the plateau lags and the target latency lags were significantly longer for PL clusters than for L clusters, indicating that the timing coordination was significantly wider in PL than in L clusters (PLAG: S1: $F[1, 87]= 15.8$, $p < 0.001$, S2: $F [1,119]= 23.8$, $p < 0.001$; TLAT: S1: $F=6.8$, $p<0.05$, S2: $F=7.6$, $p<0.01$).

3.2.2 Place of Articulation

As evident from Fig. 3, L clusters with the bilabial C_1 showed a significantly greater gesture overlap ($F[1, 39] = 13.6$, $p < 0.001$ for S1 and $F[1, 59] = 8.1$, $p < 0.01$ for S2) than those with the velar C_1. However, these clusters also showed longer constriction plateau lags ($F[1, 39] = 13.1$, $p< 0.001$ for S1 and $F[1, 59] = 12.8$, $p < 0.01$ for S2) and longer target latency lags ($F[1,39] = 17.2$, $p< 0.001$ for S1, non significant (NS) for S2), meaning that the overlap in these measurements was smaller for bilabial C_1.

Labial C_1 in sequences with reduced vowels showed a longer gesture overlap ($F[1, 46] = 28.6$, $p< 0.001$ for S1 and $F[1, 58] =37.0$, $p < 0.001$ for S2). The plateau lag was significantly longer for S1 ($F[1, 46] =41.7$, $p< 0.001$), but not significantly longer for S2. Target latency showed no significant effects.

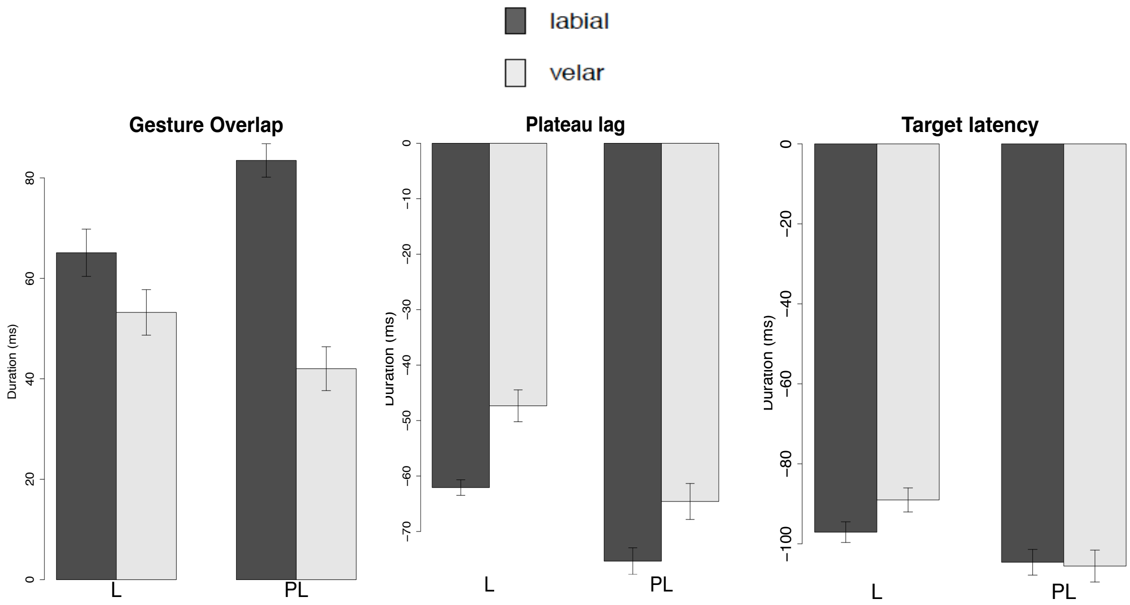

Figure 3. Normalized gesture overlap, plateau lag as well as target latency for labial and velar place of articulation of C1, averaged over two speakers and two places of articulation. A positive value means an overlapping interval of the consonant gestures and a negative value a lag between them.

3.2.2 Position in the word

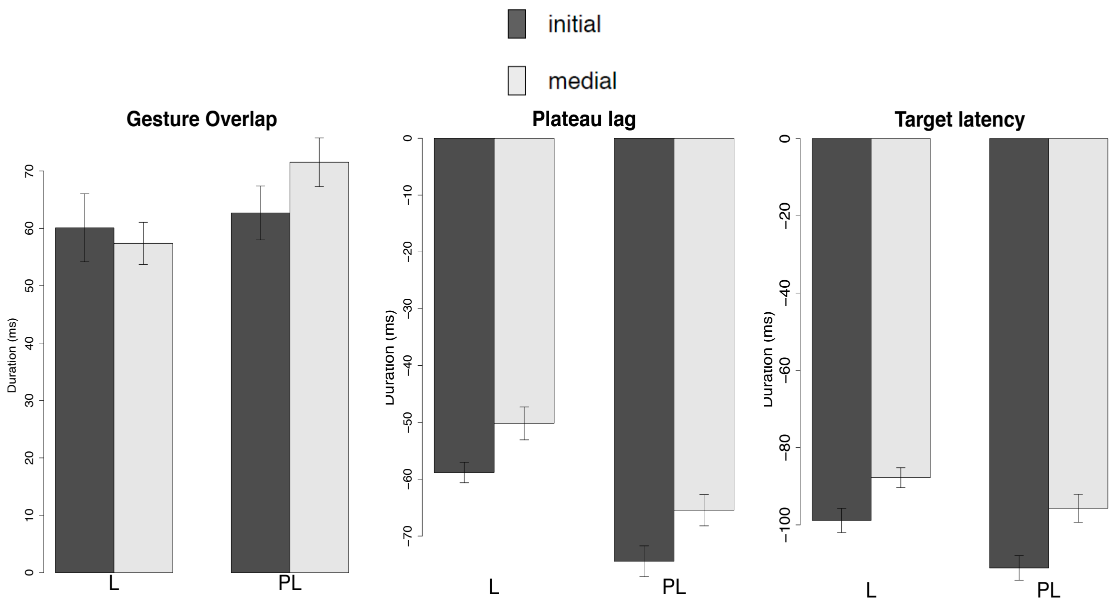

Figure 4. Gesture and plateau overlap and target latency for initial and medial word position, averaged over two speakers and two positions. A positive value means an overlapping interval of the consonant gestures and a negative value a lag between them.

Regarding cluster position in the word, L-clusters showed a slightly longer gesture overlap word initially, that was significant for S1 ($F[1, 39] = 6.4$, $p < 0.05$). Plateau lag and latency lag were both longer word initially: both were significant for S2 ($F[1, 59] = 5.5$, $p < 0.05$ and $F[1, 59] = 5.9$, $p < 0.05$), but not for S1.

There was a similar trend for PL clusters: the gesture overlap differences did not show a significant effect. Plateau lag and latency lag were both longer word initially and significant for S2 ($F[1, 58] = 4.5$, $p < 0.05$ and $F [1, 58] = 12.3$, $p < 0.001$). Onset clusters showed in both measurements longer lags than medial clusters; Consequently, gestures were timed further apart initially than medially.

4. General Discussion

This study confirmed that Portuguese PL clusters were produced with a wider phasing than their lexical counterparts. As for additional factors conditioning differences in timing, the results were mixed. Gesture overlap was greater for the bilabials, but the plateau was less overlapped for the same clusters. This may be connected with the absolute duration of the gestures and constrictions plateaus: the /p/ gesture was 40 ms longer than the /k/ gesture in L clusters and 43 ms longer in PL clusters, whereas the constriction plateau was 6 ms shorter in L /p/ than in L /k/, and 12 ms shorter in PL /p/ than PL /k/.

The gesture overlap measurements did not show a clear effect on overlap differences between places of articulation across L and PL clusters. Over the remaining measurements, the overlap between L and PL clusters did not depend on place of articulation, since it is stably greater in PL than in L over positions. Regarding position in the word, the extent of overlap between clusters was found to be greater word-medially than word-initially in both L and PL clusters, consistent with previous findings and with our predictions. The overlap differences between L and PL stayed constant also medially, where the overlap was overall greater, but it was still greater for L than PL clusters.

Overall, although L and PL clusters were segmentally similar and did not show a tongue gesture for the deleted vowel, they showed small but consistent differences in timing. These differences suggest that lexical and post-lexical clusters are not completely neutralized in EP, and EP listeners may use these subtle differences in discriminating (albeit not very well) between these cluster types.

Acknowledgments

We thank Phil Hoole for the script for the post-processing of the Ema-data and comments, Lasse Bombien for the physiological annotations' tool in EMU and all the participants of the experiments.

References

Bombien, L. Segmental and prosodic aspects in the production of consonant clusters — On the goodness of clusters. Munich: PhD-Thesis, 2011.

Browman C., & Goldstein, L. Tiers in Articulatory Phonology. In J Kingston & M. E. Beckman (Eds.), Papers in laboratory phonology 1: between the grammar and the physics of speech. Cambridge: Cambridge University Press: 241-376, 1990.

Browman, C., & Goldstein, L. Articulatory Phonology: an overview. Phonetica 49 (3-4), 155-180.1992.

Byrd, D. Influences on articulatory timing in consonant sequences. Journal of Phonetics 24: 209-244, 1996.

Chitoran, I., Goldstein, L. and Byrd, D. Gestural overlap and recoverability: Articulatory evidence from Georgian. In C. Gussenhoven and N. Warner, editors *Laboratory Phonology* VII. Berlin: Walter de Gruyter: 419-448, 2002.

Cunha, C. A elisão vocálica no Portuguêu Europeu. In B. Meisnitzer, M. Arden & C. Märzhäuser (Eds.): Tendenzen der gegenwärtigen lusophonen Sprachwissenschaft / Tendências actuais da Linguística Portuguesa, München: Maidenbauer. *in press*.

Geng, C., Mooshammer, C., Nam, H. & Hoole, P. Schwa deletion under varying prosodic conditions: Results of a pilot study. In S. Fuchs, P. Hoole, C. Mooshammer and M. Zygis (eds.), Between the regular and the particular in speech and language. Frankfurt am Main: Peter Lang:145-169, 2010

Harrington, J. The Phonetic Analysis of Speech Corpora. Wiley-Blackwell. 2010

Hoole, P., & Zierdt, A. Five-dimensional articulography. In B. Maassen & P. van Lieshout (Eds.), Speech motor control: new develpements in basic and applied resear. Oxford: Oxford University Press: 331-359. 2010.

Kühnert, B., Hoole, P. and Mooshammer, C. Gestural overlap and C-center in selected French consonant clusters. In H.C Yehia, D. Demolin & R. Laboissière (eds.) Proc. 7th International Seminar on Speech Production, UFMG Belo Horizonte: 327-334, 2006.

Marin, S. and Pouplier, M. Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control* 14 (3): 380-407, 2010

Mateus, M.H.M. and Martins, M.R.D. Contribuição para o estudo das vogais átonas no português europeu, *Biblos* 58: 111-25, 1982.

Silva, D. J. The variable deletion of unstressed vowels in Faialense Portuguese, *Language Variation* and Change 9:295–308, 1997.

Schiel, F. Automatic phonetic transcription of non-prompted speech. In Proc. 14 icphs: San Francisco. 1999.