

Portuguese Lexical Clusters and CVC Sequences in Speech Perception and Production

Conceição Cunha

Institute of Phonetics and Speech Processing, LMU Munich, Munich, Germany

Abstract

This paper investigates similarities between lexical consonant clusters and CVC sequences differing in the presence or absence of a lexical vowel in speech perception and production in two Portuguese varieties. The frequent high vowel deletion in the European variety (EP) and the realization of intervening vocalic elements between lexical clusters in Brazilian Portuguese (BP) may minimize the contrast between lexical clusters and CVC sequences in the two Portuguese varieties. In order to test this hypothesis we present a perception experiment with 72 participants and a physiological analysis of 3-dimensional movement data from 5 EP and 4 BP speakers. The perceptual results confirmed a gradual confusion of lexical clusters and CVC sequences in EP, which corresponded roughly to the gradient consonantal overlap found in production.

© 2015 S. Karger AG, Basel

1 Introduction

This study compares the production and the perception of consonant clusters that are specified in the lexicon and similar lexical CVC sequences that may be realized as phonetic clusters after high vowel deletion in unstressed syllables in European Portuguese (EP) and with a full vowel in Brazilian Portuguese (BP).

EP and BP present a similar syllable structure (Bisol, 1999; Mateus and d'Andrade, 2000; Frota and Vigário, 2001), but are postlexically affected by contradictory phonological processes such as vowel deletion and vowel epenthesis, the former being more frequent in EP and the latter in BP. After the application of these processes, the differentiation between lexical clusters and CVC sequences seems to be minimized so that EP tends to realize both as consonant sequences and BP realizes some lexical clusters with an intervening vocalic element that minimizes the difference between these sequences and sequences with a full vowel.

The work cited above, within the generative phonology framework, makes use of phonological processes such as vowel deletion and epenthesis to describe differences

between phonological and phonetic form, e.g. /felisi'dade/ realized as [fɨlisi'ðaði] and after vowel deletion and other phonological processes as [flisi'ðað] *felicidade* 'happiness' (Mateus and d'Andrade, 2000). Another plausible way of representing such differences is in terms of timing differences within the model of articulatory phonology (Browman and Goldstein, 1986, 1988), which could be expressed by a strong consonantal overlap in lexical clusters (Browman and Goldstein, 1992b) and less or no overlap of the consonants in a CVC sequence, even after vowel deletion as in *support* (Hall, 2003; Davidson, 2006). In this line of research, the segmental makeup of the consonants was shown to have an influence on the amount of overlap. Therefore, special attention will be given to the place of articulation of the consonants in the target sequences in this paper in order to test whether front-to-back clusters overlap more than clusters in the back-to-front order as previously shown by Chitoran et al. (2002).

In order to test these hypotheses, this study includes a discrimination task with minimal pairs differing in the presence of a pretonic vowel and an articulatory analysis of similar pairs in both Portuguese varieties. The next section begins with some background on the relationship between production and perception, which motivated the current experimental study. Short overviews of vowel deletion (section 1.2), vowel epenthesis (section 1.3) and distribution of rhotics (section 1.4) in the two Portuguese varieties will be followed by a quick review of place order findings in the previous literature (section 1.5). The perceptual study will be presented in section 2, followed by a physiological electromagnetic articulography (EMA) study of the same consonantal sequences in section 3. A general discussion (in section 4) bundles the results of both studies in order to discuss the relationship between production and perception, synchronic variation and sound change, and a general conclusion (in section 5) finalizes the paper.

1.1 Relationship between Production and Perception

The interaction between speaker and listener in speech communication is highly dependent on the context in which it is produced: for example, the initial consonants of *see* and *Sue* are acoustically different, because the lips of /s/ in *Sue*, but not in *see*, are already protruded in anticipation of the following vowel that is itself produced with protruded lips. In the same line, word pairs differing on the presence of a pretonic unstressed vowel as *bray* versus *beret* show small timing differences (Browman and Goldstein, 1988, 1992b), giving evidence for a different syllable structure even if the reduced vowel was acoustically absent. Speech sounds are therefore produced in a temporally overlapping way in speech communication, and there is ample evidence that this coproduction is extensive and spans much more than just neighbouring consonants and vowels (Grosvald, 2009), even crossing syllable and word boundaries.

However, fluent speakers of the language recognize that words are built out of a finite set of permutable, context-independent units that function to distinguish meaning. This abstraction allows first language adult speakers to recognize the same initial consonant in *see* and *soon*, even though there are such marked acoustic differences between them. On the other side, they are fine attuned to slight temporal differences and are expected to recover even an acoustically absent vowel in *beret*. Nonetheless, listeners differ in the way that they categorize the same coarticulated segment, because they do not factor out coarticulation perceptually to the same extent (Beddor et al., 2007; Beddor, 2009; Harrington et al., 2013). This variation in perception may be

linked to the corresponding speaker variability in production, if as argued by Fowler (2005), there is a ‘common currency’ of gestures by which coarticulation is layered in speech production and parsed in speech perception. But, if there is idiosyncrasy due to different learning experiences, then the outcome should be listener variability in parsing coarticulation of the very kind that has been found in these studies. Consequently, as Beddor et al. (2007) show, for anticipatory nasal coarticulation some listeners compensate completely for coarticulation in attributing the vowel nasalization to its source (the nasal consonant), while others only partially compensate for this effect and associate nasality to the vowel.

1.2 Vowel Deletion in Portuguese

Portuguese presents 7 contrastive stressed vowels /i, u, e, o, ε, ə, a/ (Mateus and d’Andrade, 2000), which are numerically reduced in unstressed position. In BP, the same vowel inventory in an unstressed position without the open mid vowels /ε, ə/ is assumed (Wetzels, 1992; Bisol 2005, pp. 181–185), but Barbosa and Albano (2004) argue for a strong vowel centralization and the merge of the two back vowels, resulting in the following unstressed system [ɪ, ʊ, ě, ɐ]. The unstressed vowels in EP result from a further process of vowel reduction which includes the raising of all non-high vowels and the centralization of the non-high non-back vowel, giving rise to a 4-vowel inventory with 3 high and 1 low vowel [i, ɪ, u, ɐ] (Vigário, 2003, pp. 67–69). Some of these vowels can be deleted in connected speech.

The deletion of unstressed vowels – mostly the centralized schwa in connected speech and other weakening processes such as vowel devoicing or shortening – has been reported in great detail for many languages, including English (Manuel et al., 1992; Davidson, 2006, for American English; Glowacka, 2001, for British English), German (Kohler, 1990; Nübling et al., 2006, pp. 32–34; Szczepaniak, 2007), French (Fougeron and Steriade, 1997; Torreira and Ernestus, 2011), Italian (Garrapa, 2011) and some Spanish varieties in Latin America (Boyd-Bowman, 1952; Lope Blanch, 1963; Serrano, 2006, for Mexico; Delforge, 2008a, b, 2009, for Peru; Gordon, 1980, for Bolivia; Lipski, 1990, for Ecuador).

In Portuguese, high vowel deletion has been observed for the European variety at least since 1975 (Martins, 1975). This initial study analysed 40 sentences with 20 segments from one speaker of Lisbon. It reported about 11% deletion (49 out of 394 vowels) with [u] being the most deleted vowel, followed by [ɪ] and [ɐ]. Nasal vowels and glides did not show deletion. The perceptual study of Martins and Cabral (1981) with 55 listeners reported less identification of high and mid high vowels [u, i, o] than of the low vowel [ɐ], giving some evidence for a better identification of the intrinsically longer low vowel. In order to test the identification rate of the high central and back vowels [ɪ, u], respectively, Mateus and Martins (1982) tested 18 minimal pairs differing in these vowels word finally: 188 out of 912 vowels were misidentified, meaning that 20% of the analysed vowels could not be correctly identified. Vowel quality had no influence on the results. In the same perceptual study, the authors additionally tested minimal pairs differing in syllable number. The task was to write down the number of perceived syllables for each stimulus. From all analysed minimal pairs, the pairs with the most and the least correct identifications will be reported here due to their inspiring character for the present study. The results showed some confusion of *perece* ‘spoils’ versus *prece* ‘prayer’ (5 and 4 errors out of 24, respectively). Regarding *querer* ‘to want’ versus *crer* ‘to believe’, the responses were mostly one syllable (in 20 and 23 out

of 24), meaning that the participants could not recover the vowel in *querer* any longer in perception (Mateus and Martins, 1982, pp. 179).

The corresponding vowel [i] was also the one most deleted in the production study on the island of Faial in the Azores, followed by [u] (Silva, 1997, p. 303). The lower vowel [ɐ] was mostly realized. Word position of the syllable had a significant influence in this study, with final syllables showing more deletion than non-final syllables. A further study pointed out a likelihood of deletion conditioned by stylistic and linguistic factors, in which vowel deletion marks differences in speech style. Devoicing was related to the marking of the speaker's social position in the community (Silva, 1998). A more recent study with 4 female speakers of Porto reported over 80% deletion for the high vowels [i, u], less than 10% for the unstressed vowels [i, ɐ] and no deletion of the following vowels [a, ɛ, ɔ] that can rarely occur in unstressed position due to historical reasons (Cunha, 2011, pp. 147–148).

Additionally, vowel reduction and deletion is strongly affected by word and syllable position, as well as word frequency (Vigário, 2003, chapt. 5, 7). For example, vowels in branching nuclei do not undergo vowel reduction, non-back vowels in prosodic word final position are obligatorily deleted, and some instances of vowel deletion and other specific reduction phenomena only affect a small set of highly frequent words. From the three analysed word positions (initial, medial and final) in Cunha (2011, pp. 149–150), final position attested the highest percentage of deletion and medial position the greatest variability. The reason for the great amount of variability could be the distance of the unstressed syllable relative to the main stress (as in Martins et al., 1995) and additional prosodic conditions for which further investigation is needed. In first language acquisition data, children did not produce unstressed [i] word initially before [ʃ] and did not reduce this vowel when it was the only element of the syllable (Freitas, 1997). Stressless vowels may also delete in EP (like in BP) due to hiatus resolution processes across words: for example, a word-final back vowel followed by another vowel may optionally be deleted, unless it creates a stress clash (for further details and examples, see Frota, 2000, pp. 83–85; Vigário, 2003, p. 108). The non-back central high vowel [i] is categorically deleted in prosodic word-final position, even when followed by another vowel or by a consonant giving rise to illegal consonant clusters ([ˈfortˈpsoɐ] *forte pessoa* ‘strong person’). However, the deletion is blocked, also in a categorical way, in compound-like units when that vowel is followed by a vowel bearing the compound's stress, as in [ˈgrãˈdjarjɐ] *grande area* ‘penalty area’, and in some numerals (e.g. [ˈɔzjɔɾɐ] *onze horas* ‘eleven o'clock’, Vigário, 2003, pp. 104–105).

In BP, the deletion of unstressed vowels is restricted to some segmental contexts and shows strong prosodic conditioning: high and mid vowels can be deleted before an alveolar fricative in coda ([ˈpɔts] *potes* ‘pan, pl.’) or word medially before the onset of the next syllable ([meˈdsinɐ] *medicina* ‘medicine’). This also occurs in voiced contexts ([dzaˈnovi] *dezanove* ‘nineteen’) and across word boundaries ([ˈpɔdser] *pode ser*, ‘can be’; Bisol, 1991; Bisol and Hora, 1993). In posttonic position word medially, unstressed unrounded mid vowels can be centralized ([ˈnumɐro] *número* ‘number’; Barbosa and Albano, 2004). The resulting vowels and low vowels in posttonic and prefinal position can be deleted, when the resulting cluster is legal in the Portuguese phonology (e.g. [aˈbɔbra] instead of /aˈbɔbɔra/ *abóbora* ‘pumpkin’; Bisol, 2000, 2010). Meneses and Albano (2015, this volume) showed further experimental evidence of vowel weakening, devoicing and deletion (apocope). They attested to a lower average centroid of [s]

where the vowel is not visible, longer durations of [s] and misidentification of fully devoiced vowels.

In spoken EP, vowel deletion is so extensive that consonant sequences increase massively in number and complexity: For example, [fʎisi'dad] *felicidade* 'happiness' and [ˈfrido] *ferido* 'hurt' are usual realizations in this variety, whereas in BP these forms are more commonly realized with full vowels [fɛˈlisi'dade] *felicidade* 'happiness' and [fɛˈrido] *ferido* 'hurt'. In EP, the number of deleted vowels can increase to 3 in following syllables, giving rise to consonant sequences of up to 5 or 6 consonants ([dʃprvnɪr] *desprevenir* 'fail to provide'; [dʃprtʃi'zjar] *desprestigiar* 'to depreciate'; [dzʃˈprar] *desesperar* 'to despair'). These sequences are not restricted by unusual combinations of voicing, manner and place of articulation ([tl'fɔn] *telefone* 'telephone'; [dvˈdor] *devedor* 'debtor'; [ɛ̃pdrˈnir] *empedernir* 'to petrify'; Mateus and d'Andrade, 2000, pp. 42–44).

Some of the raised phonetic clusters are similar to lexical ones, resulting in near homophones (e.g. [k(i)rɛr] *querer* 'to want' and [krɛr] *crer* 'to believe'). Studies on similar minimal pairs differing in the presence/absence of the pretonic schwa in English (e.g. beret/bray; police/please; support/sport) showed that deletion can be better explained with increased gestural overlap of both consonants than with a deletion rule, since the distinction between both words was maintained in terms of overlap: words with lexically unstressed vowels showed a longer lag between both consonants, which is consistent with the idea that consonants in lexical clusters overlap more than in sequences with an intervening lexical vowel (Browman and Goldstein, 1990, 1992b; Davidson, 2006; Geng et al., 2010). However, it is still unclear how far these tiny phonetic differences can be used by listeners to identify such word pairs differing in the presence or absence of a pretonic vowel in perception, which was the first main research question of the present study.

The resulting prediction from the cited studies is that such minimal pairs can be distinguished in terms of overlap in production, and therefore listeners should be able to recover the reduced vowel. However, the results of a previous perception experiment with similar minimal pairs in Portuguese (*crer/querer* 'to believe/to want' and *prece/perece* 'prayer/spoils') showed that EP listeners could distinguish *prece* and *perece*, but not *crer* and *querer* (Mateus and Martins, 1982, pp. 177–179). Given these results, we predict that some of the minimal pairs cannot be distinguished in perception, possibly because the amount of consonantal overlap is neutralized in production (which will be tested in the production study in section 3).

1.3 Vowel Epenthesis

A further issue of the current paper is whether an overlap account can also explain the addition of epenthetic elements. In the production of some consonant clusters, the contrary trend has been described, in which lexical clusters are realized with an intervening vocalic element (Silva, 1996), also designated *epenthetic* (Bisol, 1999; Carvalho, 2004; Collischonn, 2003; Veloso, 2007) or *svarabhaktic vowel* (Nishida, 2009). Such an element was described in both varieties and in different cluster types, as shown in the following examples from previous literature: [pilãtɛ] *planta* 'plant' in EP (Veloso, 2007) and [kapitar] *captar* 'capture' in BP (Bisol, 1999; Mateus and d'Andrade, 2000). In BP, this element has been described as being [i] or [ɪ] (Barbosa and Albano, 2004, p. 227) or [i] and [e] (Bisol, 1999). In EP, the epenthetic element was always [i] and was described as appearing in the speech of old speakers (Veloso,

2007) and of children in the early states of language acquisition (Freitas, 1997) or for prosodic reasons (Frota, 2014): when sequences of tones fall in a single stressed syllable, for example in the question *Ela foi ver o mar?* ‘Has she gone to see the sea?’, the last word *mar* bears an H + L* pitch accent and then a rise to an LH% in the boundary. For the realization of both contours, an epenthetic vowel appears at the word’s end, and the boundary tone is realized on the epenthetic vowel (Frota, 2014).

The phonological analysis of these sequences usually associates epenthetic elements with the nucleus position of a new syllable (Bisol, 1999; Mateus and d’Andrade, 2000; Veloso, 2007), meaning that lexical clusters became CVC sequences postlexically. However, an intervening vowel forming part of the word’s lexical representation may not be the only possibility of analysis. Another equally plausible way of representing such differences may be in terms of overlap: In the case that C² overlaps less with C¹ and is timed to occur later than C¹ in a /C¹C²/ cluster, this gives rise to the perception of an intervening short vowel without one actually being included in the word’s lexical representation. Some recent perceptual data attested the presence of perceptual epenthesis between illegal cluster sequences in BP (and also in Japanese), but not in EP (Dupoux et al., 2011). The second aim of this study was to test whether there is indeed a greater likelihood of perceiving vocalic intervening elements (C^VC) associated with a more separated coordination of the flanking consonants and an intervening temporal lag (i.e. wider consonantal coordination) in some intended lexical clusters in BP than in EP.

1.4 Rhotics in Portuguese

Rhotics show some phonetic variability in their realization in both Portuguese varieties; however, they contrast phonologically in intervocalic word medial position, which involves an alveolar tap in words as [‘karu] *caro* ‘expensive’ and an (alveolar or velar) trill in words as [‘karu]/[‘karu] *carro* ‘car’. In standard EP, the uvular trill co-occurs with voiced [ʁ] or the voiceless uvular fricative [χ], the alveolar trill is common in non-standard varieties of EP and in BP (Mateus and d’Andrade, 2000, p. 11, for EP; Silva, 1996; Pagan and Wertzner, 2007, for BP).

In other word positions, rhotics do not contrast phonemically. Alveolar taps occur word finally and in lexical consonant clusters (e.g. [‘mar] *mar* ‘see’ and [‘prate] *prata* ‘silver’) and trills word initially ([‘ratu] *rato* ‘mouse’). In this position the trills’ realization shows the same variability pointed out before for the intervocalic position, including alveolar and velar place of articulation as well as voiced and voiceless uvular fricatives. However, when forming part of a lexical consonant cluster, no variability has been reported for the realization of rhotics. In this position only alveolar taps have been reported (Mateus and d’Andrade, 2000, p. 11, for EP; Silva, 1996; Pagan and Wertzner, 2007, for BP). For this reason, even with contrast between tap and trills in intervocalic position, this study selected lexical words only with taps (or the so-called weak ‘r’; Silva, 1996) for a better comparison of minimal pairs differing in the presence/absence of a pretonic vowel without further cues for the distinction as in [‘kɾer] *crer* ‘to believe’ versus [‘k(i)ɾer] in EP and [ke‘ɾer] in BP *querer* ‘to want’.

1.5 Place of Articulation and Place Order

The third research question was to investigate whether the place of articulation of the first consonant (C¹) had an influence on both the amount of overlap and discriminability in perception. The perception experiment on EP showed more confusion of lexical and phonological clusters after velar (*crer/querer* ‘to believe/ to

want') than after bilabial consonants (*prece/perece* 'prayer/spoils'; Mateus and Martins, 1982, pp. 177–179). For other languages, there is some evidence for the so-called place order effect in stop clusters such that front-to-back clusters tend to overlap more than clusters in the opposite order, because consonants in front-to-back order are easier to recover in perception (Chitoran et al., 2002). Kühnert et al. (2006) found a similar effect with other cluster types, including some with liquids in the C² position, but they note that this effect coincides with independent articulators in clusters with bilabial C¹ such as /pr, pl/ compared with two related parts of the tongue in the production of /kr, kl/. Commensurately with Chitoran et al. (2002) and Kühnert et al. (2006), more overlap after bilabial clusters in /pl, pr/ is predicted than after velar clusters in /kl, kr/ for at least two reasons: (a) the independent articulators in bilabial-alveolar clusters may start simultaneously without influencing each other and (b) because of the easier recoverability of both consonants in front-to-back clusters in perception.

2 Perception Study

Recall that the main aim of the perceptual study was to test whether listeners of EP and BP can identify consonant clusters and CVC sequences with rounded and unrounded high vowels, when produced by speakers of each variety. Based on the previous studies of Mateus and Martins (1982), showing neutralization of *crer* 'to believe' and *querer* 'to want', and Dupoux et al. (2011), showing perceptual epenthesis by BP participants but not by EP ones, we predict more perceived clusters in the EP data and more CVC in the BP data.

A further aim of the perceptual study was to investigate whether the listener's variety also plays a role in the perception of these sequences. Here we predicted that clusters should be more readily perceived as singleton consonants by BP than EP listeners, if perception matches production: that is, if BP requires a wider phasing between consonants of a cluster in production, then BP listeners might also require a relatively wide phasing between consonants for them to hear a cluster rather than a singleton stop in perception. In particular, BP listeners should hear the majority of the tightly timed clusters in EP as singleton consonants. On the other hand, if perception does not follow production, no differences between the listener groups are expected.

Beyond the specifics of the variety differences, another aim was to assess whether the place order would have any influence on the stimulus identification. The last aim was to test whether stress of the target sequence has an influence on its perception. Following Harrington et al. (2013), less compensation for coarticulation is expected in unstressed syllables.

2.1 Method

The stimuli for the perception experiment were recorded as part of a physiological study (see section 3) with a multichannel DAT-device from 1 native speaker of BP and 1 of EP. The subjects were 2 female participants: the Brazilian participant is originally from São Paulo and was 25 years old at the time of the recordings, and the Portuguese one is from Porto and was 28 years old. The relevant stimuli for this study consisted of Portuguese words containing /pr/, /per/, /pur/, /kr/, /ker/ and /kur/ in initial position (table 1), read 4 times in random order from a computer screen. The speakers produced the target words embedded in two carrier sentences, one with a broad context *O Pedro*

Table 1. Recorded target words

Cluster			CVC		
[pr]	<i>prece</i>	‘prayer’	/e/	<i>perece</i>	‘spoils’
	<i>precede</i>	‘precedes’		<i>perecer</i>	‘to spoil’
			/u/	<i>pureza</i>	‘chastity’
				<i>purifica</i>	‘purifies’
[kr]	<i>crer</i>	‘to believe’	/e/	<i>querer</i>	‘to want’
	<i>credinho</i>	‘creditor’		<i>queridinho</i>	‘dear’
			/u/	<i>curado</i>	‘cured’
				<i>curador</i>	‘curator’

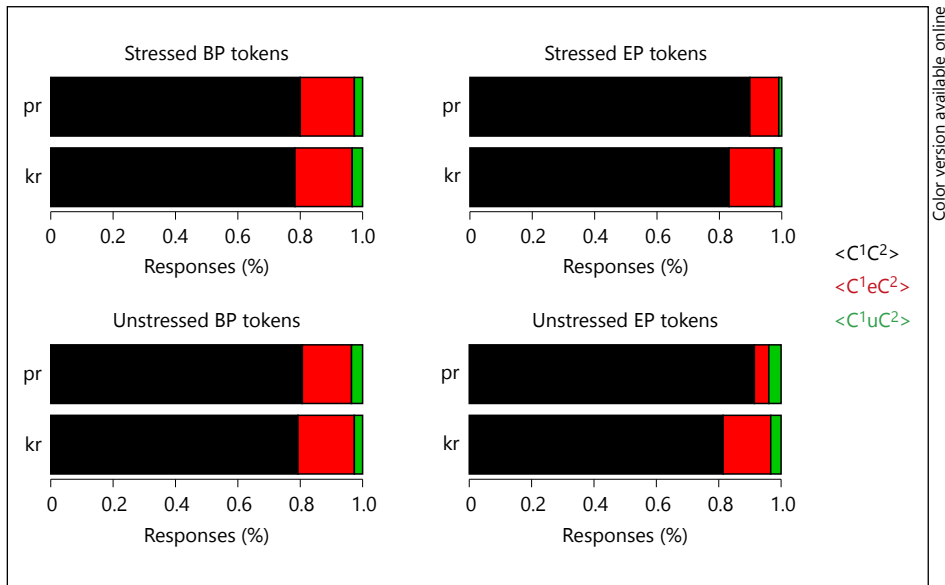
The relevant target sequences are in italics and the stressed syllables underlined.

leu ___mal ‘Peter read ___wrongly’ and the second one with a corrective focus on PEDRO: *O Tiago leu ___mal? – Não, o PEDRO leu ___mal* ‘Did Tiago read ___wrongly? – No, PETER read ___wrongly’, respectively. These prosodic differences were not analysed in the current paper.

The target stimuli from the third repetition of both speakers were excised between the beginning of the plosives’ closure and the beginning of the vocalic signal after the tap in all conditions. The resulting stimuli were 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 the following four choices <pr>, <pre>, <per>, <pur> and to each /k/-initial token with one of the following four <kr>, <kre>, <ker>, <kur>. The participants’ choices had the orthographic form they are familiar with in order to assure they would understand the task even being naïve to the purpose of the experiment.

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 statistical analysis. Thirty-seven native speakers of EP (11 males/26 females, originally from Porto and aged between 24 and 36 years) and 32 native speakers of BP (10 males/22 females, originally from Campinas and São Paulo with ages between 22 and 35 years) participated in the perception experiment. None of the subjects reported any hearing or reading problems. Although there may be contextual, dialectal and sociophonetic variation on the production of rhotics in the Portuguese varieties (Silva 1996), we selected speakers of one city in Brazil (São Paulo) and Portugal (Porto), without migration background and a similar academic profile in order to control some of this variability in this study.

The participants were subjected to both sets of stimuli. The responses were analysed with a binomial generalized linear mixed model on the binary responses (cluster vs. CVC) in the R statistical program with the listener name as a random factor and listener variety as one of the fixed factors. The listener groups will not be presented separately, because no significant influence of this variable (European vs. Brazilian listener group) was found. The statistical models will be reported in detail with the presentation of the results.



Color version available online

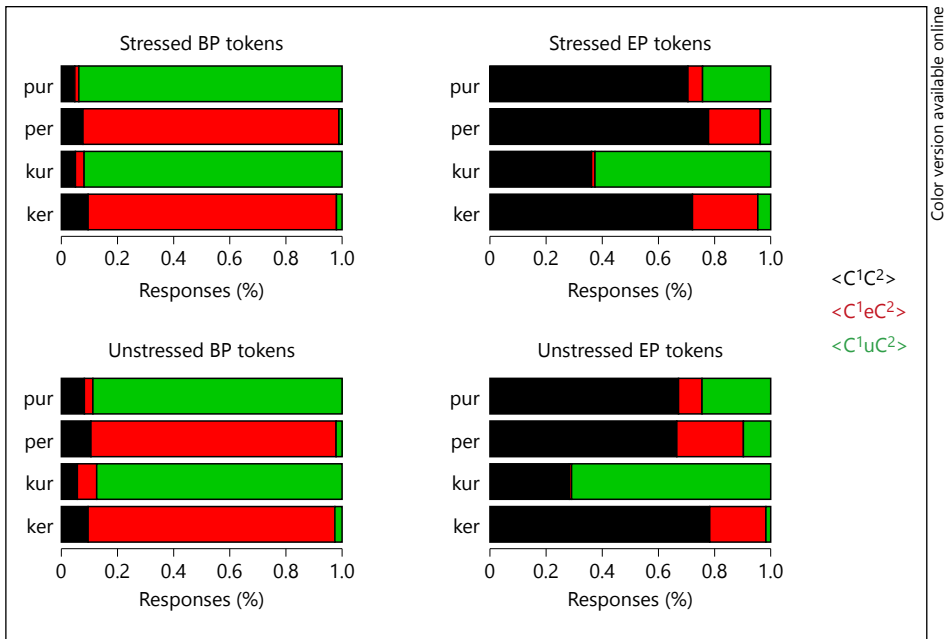
Fig. 1. Perceptual responses to the intended clusters in production, presented by stress condition (stressed vs. unstressed) for Brazilian and European production data. $\langle C^1C^2 \rangle$ corresponds to the perception of a cluster, $\langle C^1eC^2 \rangle$ to the perception of the same consonants with an unrounded vowel and $\langle C^1uC^2 \rangle$ to that of the same consonants with a rounded vowel.

2.2 Results

Figure 1 presents the perceptual responses of what were intended¹ lexical clusters in production (type *crer* ‘to believe’) – as produced by the BP speaker on the left and by the EP speaker on the right. The upper bars show lexical clusters with bilabial C^1 and the lower bars lexical clusters with velar C^1 . The target syllable was stressed in both upper panels and unstressed in the lower ones. $\langle C^1C^2 \rangle$ corresponds to the perception of two following consonants, $\langle C^1eC^2 \rangle$ to the perception of the same consonants with an intervening /e/ and $\langle C^1uC^2 \rangle$ to the same consonants with an intervening /u/. Both responses with intervening vowels were pooled for the statistical analysis.

Overall, the results showed that the presented stimuli could be overwhelmingly (>80%) identified as consonant clusters ($\langle C^1C^2 \rangle$). The mixed model on the binary response (cluster vs. sequence with vowel, irrespective of the vowel) with speaker variety, C^1 and stress as fixed factors and listener as a random factor showed a main effect of speaker variety ($\chi^2 = 4.0$, $p < 0.001$) and a significant interaction between speaker variety and C^1 ($\chi^2 = 1.9$, $p < 0.05$). Tukey post hoc t tests on the combination of speaker and C^1 confirmed what became evident from figure 2: the variety of speaker had an influence on the cluster with a bilabial C^1 but not on the one with a velar one. C^1 had an influence just on the EP judgements, but not on the BP data, meaning that

¹ Since the participants of the EMA experiment read the stimuli, it was possible to determine which word the speaker intended to realize, i.e. whether they read a word from the type *crer* ‘to believe’ or *querer* ‘to want’.



Color version available online

Fig. 2. Perceptual responses to the intended target CVC sequences in production, presented by stress condition (stressed vs. unstressed) for Brazilian and European production data. $\langle C^1C^2 \rangle$ corresponds to the perception of a cluster, $\langle C^1eC^2 \rangle$ to the perception of the same consonants with an unrounded vowel and $\langle C^1uC^2 \rangle$ to that of the same consonants with a rounded vowel.

listeners could identify more clusters after a bilabial C^1 in the EP production data, but this was not the case in the BP data, where the percentage of $\langle C^1C^2 \rangle$ was similar for both C^1 . Stress showed a small trend towards more identification of clusters in an unstressed position, which was not statistically significant.

The perceptual answers of the intended lexical CVC sequences (type *querer* ‘to want’) are presented in figure 2, arranged as in figure 1 (BP speaker on the left and EP speaker on the right, bilabial sequences in the upper two bars and the velar ones in the lower bars). In the upper panels the vowels following the liquid were stressed and in the lower panels the vowels after $/C^1/$ and $/C^2/$ were unstressed. Here both responses with vowels were also pooled for the statistical analysis.

As shown in figure 2, listeners could clearly identify the intended tokens when produced by the BP speaker, since on the left bars the perceptual answers corresponded in over 90% of the cases with the intended tokens in production. This was not the case for the EP data: with the exception of the $\langle kur \rangle$ tokens identifying words of the type *curado* ‘cured’, which could be identified in 60–70% of the cases, it is very evident that all other intended CVCs were perceived as consonant clusters in 70–80% of the occurrences. These differences explain the highly significant influence of speaker variety on the perceptual responses ($\chi^2 = 29.1$, $p < 0.001$) in the applied mixed model on the binary response (cluster vs. CVC) with speaker variety, place of articulation of C^1 and stress as fixed factors and listener as a random factor.

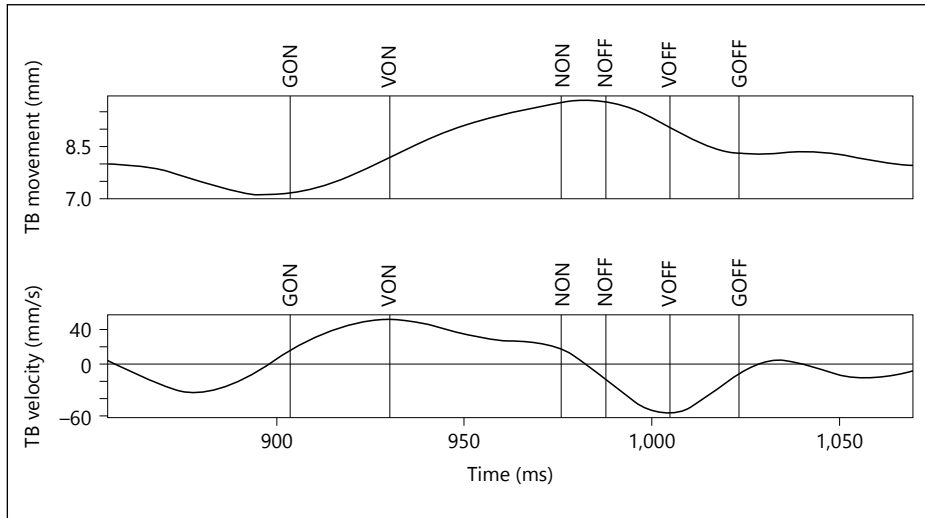


Fig. 3. Movement of tongue back (TB) in millimetres (above) and the vertical velocity of this articulator in millimetres per second (below) from /k/ in *crata* (S4, second repetition), indicating gesture onset (GON), maximum onset velocity (VON), target (NON), release (NOFF), maximum offset velocity (VOFF) and gesture offset (GOFF).

The place of articulation also had a significant influence on perception ($\chi^2 = 3.0$, $p < 0.01$) that has been restricted by the significant interaction with speaker ($\chi^2 = 4.8$, $p < 0.001$) and between speaker and C¹. Tukey post hoc t tests showed a greater influence of stress on the perception of the bilabial CVCs in EP ($p < 0.001$) than in BP ($p < 0.05$), with stressed sequences presenting less cluster perception in BP and more in EP than sequences in the unstressed condition. The velar CVCs were not influenced by stress position in any of the varieties. Place of articulation of C¹ augmented the percentage of CVC identification after a bilabial in EP to a greater extent in the stressed ($p < 0.001$) than in the unstressed condition ($p < 0.01$) and had no influence in BP. As can be seen in figure 3, the factor place of articulation augmented enormously the identification of CVCs with the rounded /u/ vowel, but not with the unrounded vowel. This factor was suppressed from the statistical analysis because it goes beyond the main questions of this paper.

Recall that listener groups (32 BP listeners vs. 37 EP listeners) were also analysed as a factor but – because they did not have an influence on the perceptual outcomes ($p = 0.2$) and for economy reasons – they were plotted together.

2.3 Interim Summary

The main global result of this perception study was that listeners of both varieties could identify cluster and CVC sequences, when produced by a BP speaker, but this distinction was less clear when produced by an EP speaker. In the latter case, listeners tend to perceive clusters irrespective of the intended token (cluster vs. CVC sequence). However, listeners still identified some of the lexical intervocalic vowels of the CVC sequences and therefore no categorical change can be assumed here and it is important to find out which factors can better model the neutralization

of both categories. Even if the differences between rounded and unrounded vowels have not been analysed, it became clear from the results that the rounded vowel could be more easily recovered than the unrounded one. Despite the big differences in the production of the tokens referred to the introduction, listener variety did not report any difference in perception.

This perceptual study replicates the findings of Mateus and Martins (1982) regarding the place of articulation of C¹, showing a stronger confusion between tokens of the types *crer* ‘to believe’ and *querer* ‘to want’, than the corresponding types with the bilabial C¹ (*prece* ‘prayer’ vs. *perece* ‘spoils’). This cannot be explained by the place order effect, which predicts more overlap in the front-to-back than in the back-to-front direction (Chitoran et al., 2002), but may be related to frequency effects. Indeed, words of the type *querer* ‘to want’ are much more frequent than *crer* ‘to believe’, and this may have favoured hypo-articulation and strong overlap. In addition, words of the type *perece* ‘spoils’ are less frequent than *prece* ‘prayer’, and this may have contributed to the hyperarticulation of the words of the second type. This study aims to test frequent words in perception, but unfortunately, due to lexical restrictions, it was not possible to match the word pairs with respect to frequency. In order to minimize the frequency effect in production in the physiological study, the set of words is changed to more infrequent words.

The percentage of identification of EP CVCs was clearly augmented by the velar place of articulation by rounded vowels and diminished in the stressed condition. At least the last finding was unexpected since there is some evidence showing more overlap in unstressed position (Byrd, 1996) and a better compensation for coarticulation in the stressed than in the unstressed condition (Harrington et al., 2013). Physiological movement data was collected in order to shed some light onto the possible articulatory grounding of these results.

3 Physiological Study

Recall that this study investigates similarities between lexical consonant clusters and lexical CVC sequences differing in the presence or absence of an unstressed vowel in speech perception and production in two Portuguese varieties. According to previous studies, the contrast between such pairs (e.g. *beret/bray*; *police/please*) were not completely neutralized, since the difference was maintained phonetically by means of a longer lag between the two consonants in the words with lexically unstressed vowels (Browman and Goldstein, 1990, 1992b; Geng et al., 2010). A similar scenario and the same trend are expected for the EP data. If the contrast cannot be maintained in terms of overlap, a neutralization of both lexical forms should be true. The rare and strongly conditioned vowel deletion in BP (Bisol, 2000, 2002) led to the prediction of a greater differentiation between lexical clusters and CVCs in BP. Another possibility would be, however, a weaker consonantal overlap and longer interconsonantal lags in BP, which could possibly explain the perceptual epenthesis found before [Dupoux et al., 2011].

We aimed to assess whether the place of articulation of the first consonant (C¹) affected the timing of lexical clusters and CVCs. Here, greater overlap is predicted in the bilabial variants than in the velar ones because of the action of independent articulators by the production of /p/ (Kühnert et al., 2006) or the place order effect

(Chitoran et al., 2002). If none of these effects apply, segmental makeup should not have an effect on gestural overlap.

Additionally, we investigated the stability of the timing as a function of stress, comparing tokens in stressed and unstressed syllables. For the CVCs we analysed sequences with primary stress in the second and in the third syllable. This means that after vowel reduction the stress position is the same in phonological and phonetic clusters. 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 C¹ (Byrd, 1996; Chitoran et al., 2002).

The last aim of the current study was to test whether the manner of articulation of the second consonant (C²) had an influence on the distinction of cluster versus CVC sequences, based on previous results showing that clusters with laterals tend to overlap more than with rhotics (Pouplier and Benus, 2012; Hoole et al., 2013). The hypothesis is that this trend is not language specific, and it should also apply to both Portuguese varieties. Although there may be dialectal and sociophonetic variation in the production of liquids in these varieties, the focus of this study is on the comparison between cluster and CVC in each variety. We expect rhotics to be produced as an alveolar tap [r] in both varieties (Mateus and d'Andrade, 2000; Silva, 1996) and the lateral approximant to be clear [l] in the Brazilian variety (Silva, 1996; Pagan and Wertzner, 2007; Wertzner et al., 2007) and dark [ɫ] in EP (Andrade, 1999; Martins et al., 2011; Oliveira et al., 2011).

3.1 Method

In order to investigate these hypotheses, physiological movement data was obtained using 3-dimensional EMA (Carstens AG500; see Hoole and Zierdt, 2010, for a description of the equipment). The physiological data was recorded at a sampling rate of 200 Hz and the synchronized audio data at a sampling rate of 32,768 Hz. 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 attached on the maxilla, nose and on the left and right mastoid bones.

3.1.1 Participants and Speech Material

The participants were 5 monolingual EP speakers from the Porto region (3 males, 2 females, and aged between 24 and 29 years) and 4 monolingual BP speakers from the city of São Paulo (3 males, 1 female, and aged between 23 and 29 years). At the time of the recordings, the speakers were exchange students or visitors in Munich and had lived there less than 10 months. They were all naïve regarding the purpose of the experiment and did not report any speech, reading or hearing disorders.

The relevant stimuli consisted of mainly lexical words presented in table 2, with the target cluster positioned word initially in the same stressed and unstressed positions as in the perception study. Due to the inexistence of lexical words of this type in Portuguese, lexical expressions were taken to fill the gap (*que lata* 'cheeky', *que latada* 'cheeky'). However, the proclitic *que* is not fully integrated within the host phonological word in both varieties (see Vigário, 2003, pp. 173–195, for EP, and Toneli, 2014, pp. 90–122; Tenani, 2002, for BP) and therefore may show even lesser overlap between the consonants than would be expected within a lexical word.

Table 2. Recorded target words for the articulatory study with English glosses

	C ²	Velar		Bilabial	
Cluster	rhotic	<i>crata</i> <i>cratada</i>	‘crater’ ‘of crater’	<i>prada</i> <i>pradada</i>	‘prada’ ‘of prada’
	lateral	<i>clata</i> <i>clatado</i>	logatome logatome	<i>plada</i> <i>platada</i>	‘surname’ logatome
CVC	rhotic	<i>queratocone</i>	‘keratoconus’	<i>perada</i>	‘pear jam’
		<i>queratoso</i>	‘with keratoconus’	<i>perador</i>	‘pear grater’
		<i>curado</i> <i>curador</i>	‘cured’ ‘curator’	<i>puratos</i> <i>puratão</i>	‘flour type’ ‘smart’
	lateral	<i>que lata</i> <i>que latada</i> <i>culatra</i> <i>culatrano</i>	‘cheeky’ ‘cheeky’ ‘breech’ ‘gun barrel’	<i>pelada</i> <i>pelador</i> <i>pulada</i> <i>pulador</i>	‘peeled’ ‘peeler’ ‘jumped’ ‘jump’

The relevant target sequences are in italics and the stressed syllables underlined.

The analysed words were embedded in the following carrier sentence *Ele leu ___ bem* (‘He read ___ right’), presented in randomized order from a computer screen. Each participant repeated the target words 8 times and had a small break after each two repetitions.

3.1.2 Data Processing

The audio signal was automatically segmented on the basis of the acoustic signal using the Munich Automatic Segmentation System (Schiel, 1999; Kisler et al., 2012; Reichel, 2012) and the target clusters were corrected manually. The physiological data was labelled with the EMU System (Harrington, 2010). The semi-automatic algorithm used could identify the peak velocities of the closing (PVEL¹) and opening (PVEL²) gesture and the maximal constriction (MAX) from the respective signal. MAX corresponds to the point of minimal velocity between the onset and offset of the target. The remaining values for gesture onset (GON) and offset (GOFF) as well as the achievement (NON) and release (NOFF) of the constriction were interpolated from the velocities’ profiles and corresponded to the 20% threshold of the difference between the closest velocities’ peaks (for a detailed description, see Bombien, 2011, pp. 71–80).

For the physiological annotation of the bilabial segments, 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 /t/ and /l/ and tongue back (tb) for /k/. Figure 3 shows an example of /k/ with the movement of tongue back in millimetres (above) and the vertical velocity of this articulator in millimetres per second (below).

3.1.3 Measurements

Overlap of the velar or bilabial C¹ and the alveolar C² was defined as the difference between the points in time from C²GON to C¹NON divided by the duration of the C¹Plateau (the plateau was defined as the duration between C¹NOFF and C¹NON) and

subtracted by 1. By subtracting it from 1, higher values indicate more overlap (Chitoran et al., 2002; Gafos et al., 2010).

- An overlap value between 0 and 1 means that C^2 starts during the C^1 Plateau, which is a common feature of cluster overlap (Gafos, 2002; Gafos et al., 2010; Chitoran et al., 2002).
- A zero or a negative value means that C^2 starts simultaneously to or after the C^1 NOFF, respectively, which means no overlap of both consonant gestures.
- A value greater than 1 means that C^2 starts before the C^1 Plateau and both consonants overlap more.

3.1.4 Statistics

The overlap variable was analysed in a mixed model (library *lme4*; Bates et al., 2011) in R (R Development Core Team, 2012) with type (3 levels: cluster, $\langle C^1eC^2 \rangle$ vs. $\langle C^1uC^2 \rangle$), speaker variety (2 levels: BP vs. EP) and C^2 (2 levels: rhotics vs. laterals) as fixed factors and speaker as a random factor. Due to the great amount of fixed factors, we carried out further analyses based on mixed models on the tokens separately for each C^2 with type (3 levels: cluster, $\langle C^1eC^2 \rangle$ vs. $\langle C^1uC^2 \rangle$), speaker variety (2 levels: BP vs. EP) and stress (2 levels: stressed vs. unstressed) and as fixed factors. In a further model, we tested the first 2 factors and C^1 . In the case of significant interactions between 2 or more factors, subsequent Tukey post hoc tests were run with the help of the *multcomp* library (Hothorn et al., 2008) in the same statistical environment.

3.2 Results

The first mixed model with C^2 , type and variety as fixed factors showed a highly significant influence of C^2 on the amount of overlap [$\chi^2(1) = 143.0$, $p < 0.001$], in which tokens with laterals showed, as predicted, more overlap in both speaker varieties than tokens with rhotics as a second consonant. For clarity reasons we further analyse the tokens with rhotics and with laterals separately.

3.2.1 Plosive and Rhotic Tokens

Figure 4 displays the overlap interval of the first consonant (C^1) with tap as a second consonant (C^2) for 4 BP (on the left) and 5 EP speakers (on the right). The upper panels present the results for the stressed tokens and the lower panels those for the unstressed tokens. As reported in the measurements above, positive values mean an overlap of the two consonants and negative values an intervening lag between the two consonants.

Consonant clusters showed smaller intervening lags than both CVC sequences in BP. This was not the case in EP, in which the overlap values of clusters and $\langle Cer \rangle$ sequences were very similar and $\langle Cur \rangle$ showed an intervening lag throughout, meaning that in these sequences both consonants did not overlap. This explains the main effects of variety [$\chi^2(1) = 7.5$, $p < 0.01$] and type [$\chi^2(2) = 55.8$, $p < 0.001$].

The place of articulation showed the expected trend of more overlap of C^2 after a bilabial than a velar C^1 [$\chi^2(1) = 7.6$, $p < 0.01$]; however, this effect has been restricted by a double interaction with variety and type [$\chi^2(7) = 65.5$, $p < 0.001$]. This trend can be seen in the EP data but was not statistically significant. The trend was inverse in the BP data with rounded vowels in $\langle pur \rangle$ ($p < 0.001$) showing shorter lags between the consonants than in $\langle kur \rangle$, but unrounded vowels showing significantly shorter lags after velars than after bilabials ($p < 0.05$). Place of articulation had no influence on clusters ($p = 1$). Despite the trend of smaller intervening distances between the consonants in BP and more overlap in

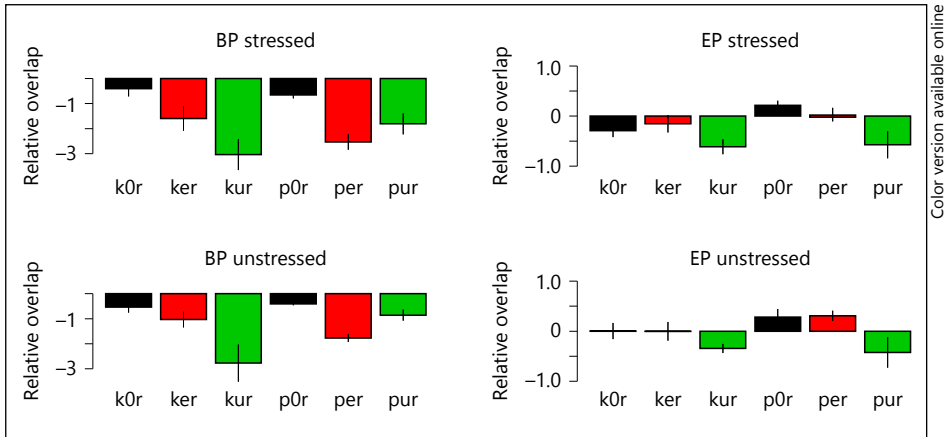


Fig. 4. Relative gesture overlap $[1 - (C^2GON - C^1NON)/(C^1NOFF - C^1NON)]$ of the tokens with plosive and rhotics. The graph shows means and standard deviations over 4 BP and 5 EP speakers. <C0r>, in which C = /p, k/, corresponds to the following clusters [pr, kr], /Cer/ and /Cur/ correspond to the CVC sequences with unrounded and rounded vowels [per, pur in BP and pir, pur in EP]², respectively. For an easier comparison, the tokens were plotted in the same grey shades (colours in the online version) as in the perception study and do not have any further meaning.

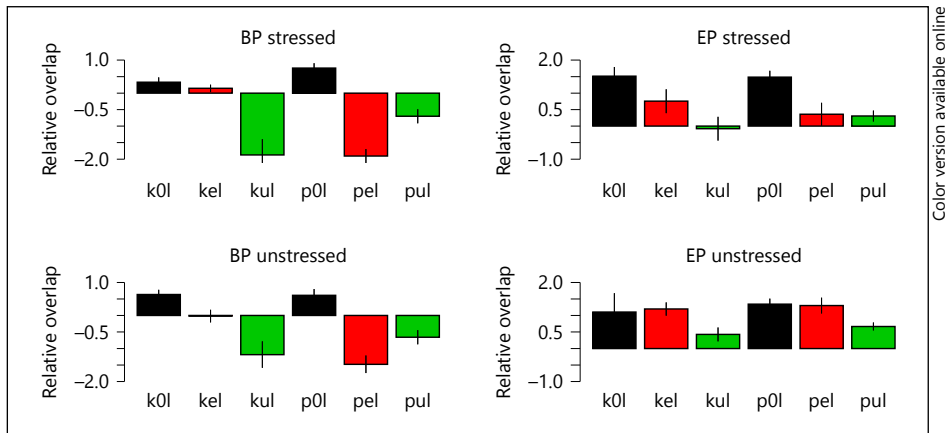
EP in unstressed than in stressed conditions, the post hoc t tests showed no influence of stress on any of the token types in both varieties. However, in the unstressed position, the overlap differences between clusters and sequences with unrounded vowels in EP were minimized, and they were no longer statistically significant.

3.2.2 Plosive and Lateral Tokens

Arranged as in the previous graph, figure 5 presents the overlap interval of C¹ with the lateral for 4 BP (on the left) and 5 EP speakers (on the right). The upper panels present the results for the stressed and the overlapping lower panels for the unstressed tokens. Recall that positive values mean overlap and negative values an intervening lag between both consonants.

The results show overlap values between both consonants in the lexical clusters and <kel> sequences in BP and lags between the consonants in all other sequences with pretonic unstressed vowels. Consequently, this overlap measurement did not attest a categorical differentiation of cluster and CVCs on the basis of positive or negative overlap. In EP, the overlap of both consonants was, as expected, even greater than in BP. All EP tokens show positive overlap values, showing at least some consonantal overlap in all three different types. These observations could be confirmed in the statistical analysis with the prominent main effect of vowel [$\chi^2(2) = 108.8, p < 0.001$] and the small influence of variety [$\chi^2(1) = 5.9, p < 0.05$]. Stress and C¹ had no significant effect on consonantal overlap, but the significant interactions can explain some of the fine-grained differences visible in figure 5.

² The orthographic versions of the token forms were chosen in all figures for internal coherence, because some of the analysed phonemes have a different realization in the two varieties (e. g. [e] in BP and [i] in EP, clear lateral in BP [l] and dark in EP [ɫ]).



Color version available online

Fig. 5. Relative gesture overlap $[1 - (C^2\text{Onset} - C^1\text{Target}) / (C^1\text{Release} - C^1\text{Target})]$ of the tokens with plosive and laterals. The graph shows means and standard deviations over 4 BP and 5 EP speakers. $\langle C0l \rangle$, in which $C = /p, k/$, corresponds to cluster, $\langle Cel \rangle$ and $\langle Cul \rangle$ correspond to the CVC sequences with unrounded and rounded vowels, respectively. These are realized as $[kel, pel]$ in BP and $[ki\text{ɫ}, pi\text{ɫ}]$ in EP for the variants with unrounded vowels and $[pul, kul]$ in BP and $[pu\text{ɫ}, ku\text{ɫ}]$ in EP for the tokens with round vowels. For an easier comparison, the tokens were plotted in the same grey shades (colours in the online version) as in the perception study and do not have any further meaning.

The place of articulation of C^1 did not have the expected effect on clusters, since consonants forming a lexical cluster did not overlap more after bilabial than after velar C^1 . However, this was clearly the case in BP in the sequences with the rounded vowels ($p < 0.001$), with the unrounded vowel velars overlap more with the bilabial ($p < 0.001$). In the EP data no effect of place of articulation was found due to the similar amount of overlap after velar and bilabial C^1 . Recall that the tokens with the velar C^1 and unrounded vowels (*que lata, que latada*, ‘cheeky’ for both) are lexical expressions and not lexical words and they have a higher boundary. In this analysis, these words showed an intervening overlap value between clusters and $\langle C^1uC^2 \rangle$ and a similar overlap pattern to the corresponding tokens with the tap $[r]$, showing less evidence for overlap differences between words and lexical expressions.

Unstressed tokens seem to be more closely coordinated than stressed ones, since they showed higher positive values, meaning more overlap and smaller lags. Overall, lexical stress also had no significant influence on overlap, but it minimized the differences in the EP data such that the overlap was not significant in both sequences with intervocalic vowels in the stressed condition. In this variety, clusters and $\langle Cel \rangle$ were also similar in the unstressed condition. There was a similar trend for BP, in which cluster and $\langle pel \rangle$ showed a similar overlap in the stressed condition, but this was not statistically significant.

4 General Discussion

The first major point of this work was the confirmation that adult hearers of EP fail to compensate for some unstressed vowels in the pretonic position as already reported some decades ago by Mateus and Martins (1982). This misperception was,

however, not categorical but continuous, and it could be better captured by some of the analysed variables with special focus on the place of articulation and vowel quality: overall, tokens with velar C¹ and central high vowels showed more confusion with lexical clusters than tokens with bilabial C¹ or back high vowels. Consequently, the lexical contrast between minimal pairs differing in the presence or absence of a lexical vowel is partially neutralized in the perception of EP tokens. In the production experiment, minimal pairs with the same segmental material presented a similar amount of interconsonantal overlap of the flanking consonants, giving ample evidence for a strong relationship between categorical perception and gestural overlap. The combined results from the perception and production experiments form a consistent picture showing that minimal pairs differing in the presence of a pretonic unstressed vowel, which were confused in perception, present a similar interconsonantal overlap in production.

Moreover, the analysed variables could describe a fine-grained variation of overlap between pairs that can be clearly distinguished in production as well as in perception (e.g. *curado* in both varieties) and neutralized pairs that could not be distinguished either in production or in perception as *crer* versus *querer* in EP. This study reported similar results for the high central vowel [i], which were very similar to the schwa in English and other languages (Davidson, 2006; Manuel et al., 2006). However, significant differences in production and perception were reported on the two vowel qualities in EP. Sequences with the lexical central high vowel showed greater consonantal overlap in production, and they were strongly neutralized in perception, whereas overlap differences between sequences with the back vowel and consonantal clusters were consistent. The latter pairs could also be recognised to a greater extent in perception. However, the back vowel is also reported as frequently deleted in previous acoustic studies (Martins et al., 1995; Silva, 1997; Cunha, 2011). Therefore, this study makes evident that acoustics are not completely synchronized with physiology, and articulatory analyses may provide consistent differences for contrasts that have already been lost in the acoustics.

Additionally, the investigation of 3-way contrast (cluster vs. /e/ vs. /u/) could add further evidence for the abstract parsing of the gradient and continuous aspect of overlap, which is very characteristic of production to the categorical perception. A more extensive experiment on timing differences of these vowels with more contexts and further factors could contribute to a better understanding of the relationship between acoustics, physiology and perception. The analysed data in this paper did not vary the prosodic context or the position of the target sequences in the word and in the sentence. It therefore does not account for the importance of prosody in consonantal coordination. However, such an analysis would go beyond the scope of this work and has to be addressed in future research. Such a future experiment should include prosodic variation, due to the important role of word and sentence prosody on vowel deletion in both varieties (Bisol, 2000; Frota, 2000; Bisol, 2002; Vigário, 2003). Word frequency had a crucial influence on the Portuguese syllable in previous work (Freitas et al., 2006) and should also be controlled in a future experiment.

Nevertheless, the results regarding clusters versus sequences with [i] were unexpected due to ample evidence for other languages that such minimal pairs can be distinguished in production in terms of overlap (Browman and Goldstein, 1990, 1992a, b; Davidson, 2006). To the best of our knowledge, there are no previous works on consonantal overlap in Portuguese, but there is solid evidence for the neutralization

of some of these minimal pairs in perception, for more than 30 years (Mateus and Martins, 1982). A possible way to understand this neutralization in both perception and production is linking synchronic variability and diachronic change (Harrington, 2012, 2014) and the perceptual model of sound change (Ohala, 1993). In a given context of spoken communication, if speaker X shows variability between forms with and without the same unstressed vowel, a context with synchronic variation between both forms coexists. In this context, listener X may recover these unstressed vowels, even when they were deleted. However, if listener X for some reason were not able to recover the deleted vowels and realize in their own production the same sequences with the deleted vowel (for example ['kɾer] for *querer* 'to want'), the forms with a deleted vowel may propagate in the spoken communication. If the new form becomes the current form for other speakers, and perhaps for the whole community (Ohala, 1993), and assimilates gradually to lexical sequences without unstressed pretonic vowels, a sound change may have taken place and the lexical contrast may be lost. The present study provides some evidence for the contrast lost between *crer* 'to believe' and *querer* 'to want', both neutralized to ['kɾer] in EP in both perception and production.

The discussed results only apply to EP, BP shows a different timing: speaker variety had the predicted influence on consonantal coordination, in which consonants in EP were indeed shown to be timed closer to each other and overlap more than in BP. The first prediction of longer lags in CVC sequences than in consonant clusters could be confirmed for the BP production data, but not for the EP data. In the latter, overlap clearly distinguished clusters from sequences with the rounded vowel (<C¹uC²>), but the contrast was more gradual between clusters and <C¹eC²>. These two types of stimuli could be better distinguished in terms of overlap in the stressed position after bilabial plosives. The amount of overlap was more similar after velar plosives in the stressed position. In the unstressed position the amount of overlap was similar for clusters and <C¹eC²>. The participants of the perception experiment could not identify the intervening vowel in more than 65% of the <C¹eC²> tokens because the similar amount of overlap between lexical clusters and <C¹eC²> may have hidden the vowel.

Moreover, the comparison of similar stimuli in perception and production makes the match between both components of speech communication very clear. Clusters and sequences with an unrounded vowel and a velar C¹ showed similar overlap values, and at the same time this word pair could not be correctly identified in perception. On the other hand, sequences with the same vowel but with the bilabial C¹ showed more overlap differences compared with the corresponding cluster, and they were less confounded with lexical clusters than the velar sequences. Consequently, although we use different tokens in perception and production and we cannot exclude frequency effects of the tokens from the perception experiment, the match of both components of speech communication could explain the almost complete lexical homophony of these word pairs due to consonantal overlap and coarticulation in production (see also Cunha, 2012, 2015).

The gradient consonantal overlap in production could be related to a gradient perception of such minimal pairs, since some evidence was found here for the relationship between consonantal coordination and overlap in production and accuracy by the tokens' identification in perception. However, in related forms as ['kɛru] *quero* 'I want', due to reasons of stress, the vowel between stop and [r] is a full mid low vowel. No stressed items were recorded for time reasons in this experiment, but the prediction for words of this type would be a wider coordination and larger lags between

both consonants that lead to full vowel perception. No evidence was found for a more or less categorical account of overlap corresponding to a categorical perception of lexical items. Furthermore, the overlap amount interacted with the segmental make-up (less overlap with rhotics than laterals) and with the position of stress. A further finding was greater consonantal overlap of tokens with laterals than with taps. This result was in accordance with preliminary results for Germanic languages (Hoole et al., 2013) and can be related to the short duration and more ballistic articulatory tongue tip movement of the tap (Zhou, 2009), which may start shortly after the production of C¹ in order to assure the short and ballistic movement for the tap production. However, the perception of these tokens is still categorical and consequently, a further interesting point here is the role of variability in the phonetic realization of this gradual overlap and how variability is accounted for in categorical perception. This work provided some evidence for the match between gradual overlap and lexical representation associated with the categorical perception for clusters with tap in the two Portuguese varieties. Other studies involving cross-linguistic data and different segmental and prosodic contexts would be needed to explore this issue further.

Another related point was the lack of evidence for the perception of intervocalic vocalic elements in lexical consonant clusters even in BP as has been attested before for this kind of clusters (Nishida, 2009). The reason for this may be regional or it may have to do with the clusters' composition. A similar experiment with medial stop clusters provided some evidence for an epenthesis due to consonantal wide coordination in BP (Cunha, 2012, 2015), which can be related to the perceptual epenthesis found in illegal clusters (Dupoux et al., 2011).

The place of articulation did not show a consistent effect on overlap, but it seems that listeners can deal less well with strongly overlapped sequences in velar than in bilabial contexts. In this study similar amounts of overlap in production caused more confusion between cluster and CVC in perception after a velar than after a bilabial consonant. Front-to-back clusters may indeed be more easily recoverable than clusters when the other way round (Chitoran et al., 2002). Stress had only a descriptive influence on the consonantal overlap in production with a trend to overlap more in unstressed than in stressed positions, but listeners appeared to be able to compensate well for the overlap in unstressed positions. Contrary to Harrington et al. (2013), we did not find any evidence for less compensation for coarticulation in unstressed positions.

Overall, the overlap in production was more categorical in BP, and this was the reason why listeners could identify the intended words in perception. In EP, the overlap was very categorical and lost its distinctiveness for some minimal pairs, which lead to the confusion of different lexical tokens in perception. This study showed that listeners cannot distinguish *crer* from *querer* anymore and could relate that to the similar amount of overlap between both consonants in both cases. The gradual account of overlap may deal better with the variability found in the production of consonant clusters, and their relationship to their perception, than a phonological analysis with a process such as vowel deletion (Mateus and d'Andrade, 2000).

The study analysed data from only one city of each variety (São Paulo/Campinas region for BP and Porto for EP). Consequently, this study cannot sufficiently access regional differences. Furthermore, the data for the perception experiment included only one speaker of each variety and it excluded interspeaker variability completely. The data could have shown even less differences between clusters and CVC sequences if

the materials included some more natural speech given that deletion and epenthesis are more common in connected speech. However, it is currently difficult to obtain EMA data from spontaneous speech for two main reasons: it would be necessary to have two EMA systems in a room, and the complex experimental set-up inhibits casual natural speech.

5 Conclusion

Overall, the consonantal overlap in production was more categorical in BP, and it led listeners to identify the intended words in perception. In EP, the overlap was gradual, and it lost its distinctiveness for some minimal pairs, which led to the confusion of the same lexical tokens in perception. This study showed that listeners cannot distinguish *crer* ‘to believe’ from *querer* ‘to want’ anymore and could relate that to the similar amount of overlap between both consonants in both cases. At this stage, some lexical distinctive sequences are neutralized in the spoken EP in perception as well as in the production of native speakers. Presently, this is still a case of language variation, since native speakers still know both words and – with context – may distinguish them. However, since listeners cannot recover some unstressed vowels from the production of contemporary talkers any longer, one may face a mismatch between the intended production of [ˈkɪrɐ] (*querer* ‘to want’) and the perceived token [ˈkrɐr] as a reduced form of *querer* or the other lexical competitor *crer* ‘to believe’. If the listener fails to compensate for the reduced vowel and assume the reduced form [ˈkrɐr] as being the realization of *querer*, we are dealing with a sound change in progress in the sense of Ohala (1993) and Harrington (2006). Such sound changes may lead to the complete merger of both forms and consequently to greater differences between the two varieties of Portuguese. It may lead to some difficulties in the mutual intelligibility of the varieties. However, we did not find any evidence for the latter, since the factor listener group did not influence the perceptual results: EP reduced tokens were as hard to perceive for BP listeners as for native EP listeners.

A similar process of distinction lost may be responsible for the lexicalization of actual consonant clusters from earlier Latin CVC sequences. Diachronically, Classical Latin (CL) ē, ĭ and ě were deleted before the evolution to contemporary Portuguese and consequently, CL *bĕrillum* became *brilho* ‘brightness’ in Portuguese, CL *quĕrītāre* > port. *gritar* ‘shout’ and CL *sŭpĕrāre* > port. *sobrar* ‘remain’ (Boyd-Bowman, 1980, pp. 122–123; Williams, 1968, pp. 56–57, 224–233).

In a nutshell, the present study has shown how the gradual consonantal overlap and its reinterpretation with respect to perception can account for variation on contemporary productions, but also be related to the path of sound change and language evolution.

Acknowledgements

This work was partially supported by the Linguistics International Doctoral Program (LIPP) and the Graduate Center of the Ludwig Maximilian University Munich. I also thank Marina Vigário, Rachel Walker and two anonymous reviewers for thoughtful comments as well as the audience of PaPI 2013 and all participants of the experiments discussed in the paper.

References

- Andrade A (1999): On /l/ velarization in European Portuguese. *Proceedings of the International Conference of Phonetic Sciences (ICPhS)*, San Francisco, pp 543–546.
- Barbosa P, Albano E (2004): Brazilian Portuguese. *J Int Phonet Assoc* 34:227–232.
- Bates D, Maechler M, Bolker B (2011): lme4: linear mixed-effects models using Eigen and Eigen. R package version 0.999375-42. <http://CRAN.R-project.org/package=lme4>.
- Beddor P (2009): A coarticulatory path to sound change. *Language* 85:785–821.
- Beddor P, Brasher A, Narayan C (2007): Applying perceptual methods to phonetic variation and sound change; in Solé MJ, Beddor PS, Ohala M (eds): *Experimental Approaches to Phonology*. Oxford, Oxford University Press, pp 127–143.
- Bisol L (1991): Palatalization and its variable restriction. *Int J Sociol Lang* 89:107–124.
- Bisol L (1999): A sílaba e seus constituintes; in Neves MHM (ed): *Gramática do Português Falado 7*. Campinas, Humanitas/Unicamp, pp 701–741.
- Bisol L (2000): A elisão, uma regra variável. *Letras de Hoje* 35:319–330.
- Bisol L (2002): A degeminação e a elisão no VARSUL; in Bisol L, Brescancini C (eds): *Fonologia e Variação: recortes do português brasileiro*. Porto Alegre, Edipucrs, pp 231–250.
- Bisol L (2005): Introdução a estudos de fonologia do português brasileiro. Porto Alegre, Edipucrs.
- Bisol L, Hora D (1993): Palatalização da oclusiva dental e fonologia lexical. *Letras* 5:25–40.
- Bombien L (2011): *Segmental and Prosodic Aspects in the Production of Consonant Clusters*; doctoral dissertation, Ludwig-Maximilians-Universität München.
- Boyd-Bowman P (1952): La pérdida de vocales átonas en la altiplanicie Mexicana. *Nuev Rev Filol Hisp* 6:138–140.
- Boyd-Bowman P (1980): *From Latin to Romance in Sound Charts*. Washington, Georgetown University Press.
- Browman CP, Goldstein L (1986): Towards an articulatory phonology. *Phonol Yearb* 3:9–252.
- Browman CP, Goldstein L (1988): Some notes on syllable structure in articulatory phonology. *Phonetica* 45:140–155.
- Browman CP, Goldstein L (1990): ‘Targetless’ schwa: an articulatory analysis. *Haskins Laboratories Status Report on Speech Research*. Bd SR-101/102.
- Browman CP, Goldstein L (1992a): Articulatory phonology: an overview. *Phonetica* 49:155–180.
- Browman C, Goldstein L (1992b): ‘Targetless’ schwa: an articulatory analysis; in Docherty GJ, Ladd DR (eds): *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge, Cambridge University Press, pp 26–67.
- Browman C, Goldstein L (2000): Competing constraints on intergestural coordination and self-organization of phonological structures. *Bull Commun Parlée* 5:25–34.
- Byrd D (1996): Influences on articulatory timing in consonant sequences. *J Phonet* 24:209–244.
- Carvalho KC (2004): *Descrição fonético-acústica das vibrantes no português e no espanhol*. Assis, UNESP.
- Chitoran I, Goldstein L, Byrd D (2002): Gestural overlap and recoverability: articulatory evidence from Georgian; in Gussenhoven C, Warner N (eds): *Laboratory Phonology VII*. Berlin, de Gruyter, pp 419–448.
- Collischonn G (2003): Epêntese Vocálica no Português do Sul do Brasil: Variáveis Extra-linguísticas. *Rev Letras* 61:285–297.
- Cunha C (2011): A elisão vocálica no Português Europeu; in Meisnitzer B, Arden M, Märzhäuser C (eds): *Tendenzen der gegenwärtigen lusophonen Sprachwissenschaft/Tendências actuais da Linguística Portuguesa*. München, Maidenbauer, pp 133–148.
- Cunha C (2012): *Die Organisation von Konsonantenclustern und CVC-Sequenzen in zwei portugiesischen Varietäten*; doctoral dissertation, Ludwig-Maximilians-Universität München. <http://nbn-resolving.de/urn:nbn:de:bvb:19-183346>.
- Cunha C (2015): captar – cap(i)tar, facto – fato: Variation und Wandel bei Plosivsequenzen im Portugiesischen. *Promptus – Würzburger Beiträge zur Romanistik*, pp 57–87.
- Davidson L (2006): Schwa elision in fast speech: segmental deletion or gestural overlap? *Phonetica* 63:79–112.
- Delforge AM (2008a): Unstressed vowel reduction in Andean Spanish; in Colantoni L, Steele L (eds): *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*. Somerville, Cascadilla Proceedings Project, pp 107–124.
- Delforge AM (2008b): Gestural alignment constraints and unstressed vowel devoicing in Andean Spanish; in Chang CB, Haynie HJ (eds): *Proceedings of the 26th West Coast Conference on Formal Linguistics*. Somerville, Cascadilla Proceedings Project, pp 147–155.
- Delforge AM (2009): *The Rise and Fall of Unstressed Vowel Reduction in the Spanish of Cusco, Peru: A Sociophonetic Study*, doctoral dissertation, Davis: University of California.
- Dupoux E, Parlato E, Frota S, Hirose Y, Peperkamp S (2011): Where do illusory vowels come from? *J Mem Lang* 64:199–210.
- Fougeron C, Steriade D (1997): Does deletion of French schwa lead to neutralization of lexical distinctions? *Proceedings of the 5th European Conference on Speech Communication and Technology*, Rhodes, vol 7, pp 943–937.
- Fowler C (2005): Parsing coarticulated speech in perception. *J Phonet* 33:199–213.
- Freitas MJ (1997): *Aquisição da Estrutura Silábica do Português Europeu*; doctoral dissertation, FLUL, Lisboa.

- Freitas MJ, Frota S, Vigário M, Martins F (2006): Efeitos prosódicos e efeitos de frequência no desenvolvimento silábico em Português Europeu; in Duarte I, Leiria I (eds): XXI Encontro da Associação Portuguesa de Linguística. Textos Seleccionados. Lisboa, APL, pp 397–412.
- Frota S (2000): Prosody and Focus in European Portuguese. *Phonological Phrasing and Intonation*. New York, Garland Publishing.
- Frota S (2014): The intonational phonology of European Portuguese; in Jun SA (ed): *Prosodic Typology II*. Oxford, Oxford University Press, pp 6–42.
- Frota S, Vigário M (2001): On the correlates of rhythmic distinctions: the European Portuguese/Brazilian Portuguese case. *Probus* 13:247–275.
- Gafos A (2002): A grammar of gestural coordination. *Nat Lang Linguist Theory* 20:269–337.
- Gafos A, Hoole P, Roon D, Zeroual C (2010): Variation in overlap and phonological grammar in Moroccan Arabic clusters; in Cohn AC (ed): *Laboratory Phonology X*. Berlin, Mouton de Gruyter.
- Garrapa L (2011): *Vowel Deletion in Florentine Italian*. Bern, Lang.
- Geng C, Mooshammer C, Nam H, Hoole P (2010): Schwa deletion under varying prosodic conditions: results of a pilot study; in Hoole P, Mooshammer C, Zygis M (eds): *Between the Regular and the Particular in Speech and Language*. Frankfurt am Main, Lang, pp 145–169.
- Glowacka D (2001): Unstressed vowel deletion and new consonant clusters in English. *Poznan Stud Contemp Linguist* 37:71–94.
- Goldstein L, Pouplier M (2014): The temporal organization of speech; in Goldrick M, Ferreira V, Miozzo M (eds): *The Oxford Handbook of Language Production*. Oxford, Oxford University Press, pp 210–227.
- Gordon A (1980): Notas sobre la fonética del castellano en Bolivia; in Gordon A, Rugg E (eds): *Actas del sexto congreso internacional de Hispanistas*. Toronto, University of Toronto, pp 349–352.
- Grosvald M (2009): Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *J Phonet* 37:173–188.
- Hall N (2003): *Gestures and Segments: Vowel Intrusion as Overlap*; doctoral dissertation, University of Massachusetts at Amherst.
- Harrington J (2006): An acoustic analysis of ‘happy-tensing’ in the Queen’s Christmas broadcasts. *J Phonet* 34:439–457.
- Harrington J (2010): *Phonetic Analysis of Speech Corpora*. Chichester, Blackwell.
- Harrington J (2012): The relationship between synchronic variation and diachronic change; in Cohn AC, Fougeron C, Huffman M (eds): *Handbook of Laboratory Phonology*. Oxford, Oxford University Press, pp 321–332.
- Harrington J (2014): Variability and change in spoken language communication; in Glaser E, Kolmer A, Meyer M, Stark E (eds): *Sprache(n) verstehen*. Zürich, Hochschulverlag ETH Zürich, pp 33–57.
- Harrington J, Kleber F, Reubold U (2013): The effect of prosodic weakening on the production and perception of trans-consonantal vowel coarticulation in German. *J Acoust Soc Am* 134:551–561.
- Hoole P, Pouplier M, Benus S, Bombien L (2013): Articulatory coordination in obstruent-sonorant clusters and syllabic consonants: data and modelling; in Spreafico L, Vietti A (eds): *Proceedings of Raticas 3*. Bolzano, Bolzano University Press, pp 79–94.
- Hoole P, Zierdt A (2010): Five-dimensional articulography; in Maassen B, van Lieshout PHHM (eds): *Speech Motor Control: New Developments in Basic and Applied Research*. Oxford, Oxford University Press, pp 331–349.
- Hothorn T, Bretz F, Westfall P (2008): Simultaneous inference in general parametric models. *Biometr J* 50:346–363.
- Kisler T, Schiel F, Sloetjes H (2012): Signal processing via web services: the use case WebMAUS. *Proceedings Digital Humanities, Hamburg, 2012*, pp 30–34.
- Kohler KJ (1990): Macro and micro F0 in the synthesis of intonation. *Papers Lab Phonol I*:115–138.
- Kühnert B, Hoole P, Mooshammer C (2006): Gestural overlap and C-center in selected French consonant clusters; in Yehia HC, Demolin D, Laboissière R (eds): *Proc 7th International Seminar on Speech Production*. Belo Horizonte, UFMG, pp 327–334.
- Lipski JM (1990): Aspects of Ecuadorian vowel reduction. *Hisp Linguist* 1:1–19.
- Lope Blanch JM (1963): En torno a las vocales caedizas del español mexicano. *Nuev Rev Filol Hisp* 17:1–19.
- Manuel S, Shattuck-Hufnagel S, Huffman S, Stevens K, Hunnicut S (1992): Studies of vowel and consonant reduction; in Ohala J, Nearey TM, Derwing BK, et al (eds): *Proceedings of the 1992 International Conference on Spoken Language Processing*. Edmonton, University of Alberta, vol 2, pp 943–946.
- Marin S, Pouplier M (2010): Temporal organization of complex onsets and codas in American English: testing the predictions of a gestural coupling model. *Motor Control* 14:380–407.
- Martins MRD (1975): Vogais e Consoantes do Português: Estatística do Ocorrência, Duração e Intensidade. *Bol Filol* 24:1–11.
- Martins MRD, Cabral VR (1981): Percepção de consoantes e vogais do Português. *Communication du 2ème Symposium Luso-Espagnol d’Acoustique*, Madrid.
- Martins MRD, Poch-Olivé D, Harmegnies B (1995): Processos de mudança linguística em curso no português europeu; in Martins MRD (ed): *Fonética do Português: Trinta anos de investigação (2002)*. Lisboa, Caminho, pp 295–300.
- Martins P, Oliveira C, Ferreira C, Silva A, Teixeira A (2011): 3D MRI and semi-automatic segmentation techniques applied to the study of European Portuguese lateral sounds. *Proceedings 9th International Seminar on Speech Production, Montreal*, pp 235–239.

- Mateus MH, d'Andrade E (2000): *The Phonology of Portuguese*. Oxford: Oxford University Press.
- Mateus MH, Martins MRD (1982): Contribuição para o Estudo das Vogais Átonas [ə] e [u] no Português Europeu. *Biblos* 58:111–128.
- Meneses FO, Albano E (2015): From devoicing to apocope: an acoustic study of post-stressed high vowel lenition in Brazilian Portuguese. *Phonetica* 72:121–137.
- Nishida G (2009): Análise acústica do Tap em Grupos no PB. Curitiba, Universidade Federal do Paraná.
- Nübling D, Dammel A, Duke J, Szczepaniak R (2006): *Historische Sprachwissenschaft des Deutschen: Eine Einführung in die Prinzipien des Sprachwandels*. Tübingen, Narr-Verlag.
- Ohala J (1992): Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Papers from the Parasession on the Syllable*. Chicago, Chicago Linguistic Society, pp 319–338.
- Ohala J (1993): Sound change as nature's speech perception experiment. *Speech Commun* 28:155–161.
- Oliveira C, Martins P, Teixeira A (2011): An articulatory and acoustic study of the European /l/. *Proceedings of the 17th ICPhS, Hong Kong*.
- Pagan LO, Wertzner HF (2007): Análise Acústica das Consoantes Líquidas do Português Brasileiro em crianças com e sem Transtorno Fonológico. *Rev Soc Bras Fonoaudiol* 12:106–113.
- Pouplier M, Beňuš Š (2011): On the phonetic status of syllabic consonants: evidence from Slovak. *J Lab Phonol* 2:243–273.
- R Development Core Team (2012): *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Reichel UD (2012): Perma and Balloon: tools for string alignment and text processing, *Proceedings Interspeech*, Portland, p 346.
- Schiel F (1999): Automatic Phonetic Transcription of Non-Prompted Speech, *Proceedings of the ICPhS 1999*, San Francisco, August 1999, pp 607–610.
- Serrano J (2006): En torno a las vocales caedizas del español mexicano: una aproximación sociolingüística; in Butragueño PM (ed): *Líderes lingüístico. Estudios de variación y cambio*. México, El Colegio de México, pp 37–59.
- Silva A (1996): *Para a Descrição Fonético-Acústica das Líquidas no Português Brasileiro: Dados de um Informante Paulistano*; doctoral dissertation, UNICAMP/IEL, Campinas.
- Silva DJ (1997): The variable deletion of unstressed vowels in Faialense Portuguese. *Lang Variation Change* 9:295–308.
- Silva DJ (1998): Vowel lenition in São Miguel Portuguese. *Hispania* 81:166–178.
- Szczepaniak R (2007): *Der phonologisch-typologische Wandel des Deutschen von einer Silben- zu einer Wortsprache*. Berlin, Mouton de Gruyter.
- Tenani LE (2002): *Domínios prosódicos no Português*; doctoral dissertation, Universidade Estadual de Campinas.
- Toneli P (2014): *A Palavra Prosódica no Português Brasileiro*; doctoral dissertation, Universidade Estadual de Campinas.
- Torreira F, Ernestus M (2011): Vowel elision in casual French: the case of vowel /e/ in the word c'était. *J Phonet* 39:50–58.
- Veloso J (2007): *Da influência do conhecimento ortográfico sobre o conhecimento fonológico*. Munich, Lincom.
- Vigário M (1998). *Elisão da Vogal Não-Recuada Final e a Palavra Prosódica no Português Europeu*; in Lopes AC, Martins C (eds): *XIII Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa, APL, pp 359–376.
- Vigário M (2003): *The Prosodic Word in European Portuguese*. Berlin, Mouton de Gruyter.
- Wertzner HF, Pagan-Neves LO, Castro MM (2007): Acoustic Analysis and Stimulability Index for Brazilian Portuguese Liquid Sounds. *Rev CEFAC* 9:339–350.
- Wetzels WL (1992): Mid vowel neutralization in Brazilian Portuguese. *Cad Estud Lingüíst* 23:19–55.
- Williams EB (1968): *From Latin to Portuguese*. Philadelphia, University of Pennsylvania Press.
- Zhou X (2009): *An MRI-Based Articulatory and Acoustic Study of American English Liquid Sounds /r/ and /l/*; doctoral dissertation, University of Maryland, College Park.