GESTURAL COORDINATION OF BRAZILIAN PORTUGUESE NASAL VOWELS IN CV SYLLABLES: A REAL-TIME MRI STUDY

Alexsandro Meireles[1], Louis Goldstein[2], Shrikanth Narayanan[2], and Reed Blaylock[2]

[1] Federal University of Espírito Santo, Brazil
[2] University of Southern California, USA
meirelesalex@gmail.com

ABSTRACT

The coordination of velum and lingual gestures for Brazilian Portuguese nasal vowels in CV syllables is studied with real-time magnetic resonance imaging (rt-MRI). Four syllable structures were compared: stop-oral vowel, stop-nasal vowel, nasal-oral vowel, nasal-nasal vowel. Results show near-synchrony timing (in-phase) for the consonant onset and nasal vowel's lingual onset, and for the consonant release and velum onset in the stop-nasal vowel syllable structure. The other syllabic structures present near-synchrony timing (in-phase) for all their gesture's onsets.

Keywords: gestural coordination, nasal vowels, Brazilian Portuguese, articulation, real-time MRI.

1. INTRODUCTION

A recent model [4, 5] proposes that languages' syllable structure emerges from stable patterns of coordination among speech gestures. In this model, each speech gesture is associated with a non-linear planning oscillator "that triggers the production of that gesture during speech production" [5, p. 242]. The coupling of these oscillators with one another are represented by coupling graphs that represent syllable structure.

According to the *Coupling Hypothesis of Syllable Structure* (henceforth CHSS) [4, p, 232], CV syllables have the following characteristics: a) in-phase coordination of C gestures with the V gesture; b) spontaneous emergence of coordination pattern in development; and c) free combination of C and V within the grammar of languages. On the other hand, VC syllables can be characterized by a) anti-phase coordination of the V gesture and the C gesture; b) possible learning requirement with later emergence of pattern; and c) more restricted combinations.

Löfqvist and Gracco [12] have presented data showing that the C and V gestures in CV syllables are initiated synchronously, consistent with the hypothesis of the in-phase coupling. It is tempting to extend this idea to the hypothesis that all gestures in any CV syllable are initiated synchronously. However, syllables containing nasal vowels may provide evidence contrary to this stronger hypothesis. In order to differentiate [pã] from [mã], for example, it would seem that the velum lowering gesture in [pã] cannot be initiated synchronously with the C and V gestures. Compatibly, the distinction between oral and nasal vowels is acquired relatively late in the language acquisition process [7, p. 72], so that CV syllables where V is nasal probably do not emerge spontaneously. Therefore, it seems that not only within-onset and within-coda coordination may require coupling to be learned [4], but also the non-synchronous within-V coordination of the velum lowering and vowel constriction gestures.

1.1. Previous articulatory studies on Portuguese

Although there are, as far as we know, no studies on the gestural coordination in Portuguese, some research sheds some light on the subject.

The MRI Study on Brazilian Portuguese in Medeiros and Demolin [15] supports the findings of Maeda [14] and Demolin and colleagues [2] that the tongue body is more to the back for nasal vowels. Also, maybe due to the limitation of their data acquisition rate (5 frames per second), they observed velum aperture during the whole course of the nasal vowel.

Oliveira [18, p. 238] reports in her EMA study on European Portuguese that the velum is closed (or lightly opened) at the beginning of nasal vowel, which acoustically results in an oral-like vowel. This same result has been found for Brazilian Portuguese [15, 22, 20] and other studies on European Portuguese [3, 9, 8, 21]. Montagu [16] also found a late onset of velum opening in data on French nasal vowels.

According to Oliveira's MRI study on European Portuguese nasal vowels [18], although there are no measurements of the coordination between velum and tongue gestures, the data suggests that the oral constriction starts simultaneously with the stop onset, and the velum opening starts at the stop release.

Table 1: Stimuli used in the MRI experiment. Besides the 2 prosodic conditions, different carrier sentences were used in order to obtain an alternate tongue movement before and after the target vowel (high to low vowel; low to high vowel). Bold words represented the words that have been analyzed. Note that the "n" in the words does not represent a nasal consonant in Portuguese; it is used in orthography to represent nasal vowels. So, a word like "panto", for example, must be interpreted as [pã.tʊ].

Strong prosodic boundary	Weak prosodic boundary	Word type	Vowel type	Target Syllable
Acho que ouvi **pato**.	Acho que ouvi **pato** agora mesmo.	CV	a	[pa]
Acho que ouvi **panto**.	Acho que ouvi **panto** agora mesmo.	CVN	a	[pã]
Acho que ouvi **mato**.	Acho que ouvi **mato** agora mesmo.	NV	a	[ma]
Acho que ouvi **manto**.	Acho que ouvi **manto** agora mesmo.	NVN	a	[mã]
Peço que ouça **pita**.	Peço que ouça **pita** agora mesmo.	CV	i	[pi]
Peço que ouça **pinta**.	Peço que ouça **pinta** agora mesmo.	CVN	i	[pĩ]
Peço que ouça **mita**.	Peço que ouça **mita** agora mesmo.	NV	i	[mi]
Peço que ouça **minta**.	Peço que ouça **minta** agora mesmo.	NVN	i	[mĩ]
Peço que ouça **buda**.	Peço que ouça **buda** agora mesmo.	CV	u	[bu]
Peço que ouça **bunda**.	Peço que ouça **bunda** agora mesmo.	CVN	u	[bũ]
Peço que ouça **muda**.	Peço que ouça **muda** agora mesmo.	NV	u	[mu]
Peço que ouça **munda**.	Peço que ouça **munda** agora mesmo.	NVN	u	[mũ]

1.2. Motivation and Hypotheses

The motivation for the study is the lack of detailed account of the articulatory coordination of Brazilian Portuguese nasal vowels in CV syllables. Also, we will test how to extend the CHSS to a complex multi-gesture CṼ syllable.

Based on the considerations discussed above, we propose the following hypotheses regarding the phase relations in CV syllables with and without nasalization:

H1. For CVN syllables, the oral consonant-nasal vowel coordination is composed of synchronous coordination between the oral consonant onset and the vowel's oral gesture onset, non-synchronous coordination between the oral consonant and the velum lowering, and non-synchronous coordination between the vowel's oral gesture and the velum lowering.

H2. Other syllable structures such as CV, NV, and NVN will have in-phase coordination in all pairwise combinations.

2. METHODS

Minimal pairs contrasting oral and nasal vowels in CV syllables, and inserted in a carrier sentence, were read by 2 native speakers of Brazilian Portuguese (see Table 1). Each sentence was read 10 times and presented in random order. Two varieties of the carrier sentence were used in order to verify whether prosodic position had any effect on CV coordination. Subject 1 is a male speaker and one of the authors of this study, and Subject 2 is a female speaker. The subjects reported no speech or hearing pathology and were paid for their participation. Both speakers are from Conselheiro Lafaiete, Minas Gerais state (Southeastern Brazil) and have college-level education.

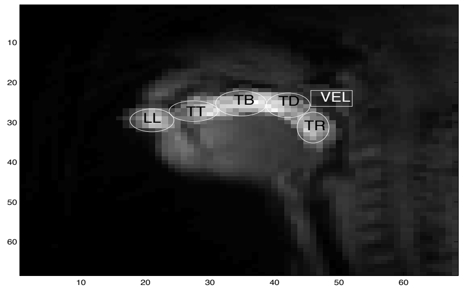
2.1. MRI Imaging

The speech data were acquired using a rtMRI protocol developed specifically for the study of speech production [17]. Subjects' upper airways were imaged midsagittally with spatial resolution 68 x 68 pixels, field of view 200 x 200 mm, and a temporal reconstruction rate of 33.18 frames per second (fps). Speech was recorded inside the scanner at 20 kHz, simultaneously with the MRI acquisition, and later noise-reduced [1].

2.2. Articulatory and Statistical Analysis

Time functions of the constrictions of the lips, tongue, and velum were estimated using changes in the average pixel intensity in speaker-specific regions-of-interest of the vocal tract (ROIs: lips [p, b, m], tongue tip [t], tongue body [i], tongue dorsum [u], tongue root [a], velum [m, ã, ĩ, ũ]) [10, 11]. Tongue tip was estimated but not considered for this study. ROIs for the lips (LL), tongue tip (TT), tongue body (TB), tongue dorsum (TD), and tongue root (TR) were measured using a set of subregions. Each subregion had a horizontal width of only one pixel, with a vertical span of four pixels beginning at the palate. These five subregions were selected from a midline that corresponds to the mean pixel intensity from the vocal folds to the lips. The velum was measured using a geometrical figure that captures its displacement along the vertical and horizontal axes. ROI locations are displayed in figure 1. To minimize noise or random intensity fluctuations, all resulting signals were smoothed by a locally weighted linear regression [10]. The weighting function used was a Gaussian kernel K with a standard deviation of h samples, where h = .8, giving a smoothing window width of roughly 90 ms given the sampling period of 30.14 ms.

Figure 1: Regions of interest (ROIs) for the articulatory analysis.



MRI data were loaded into a MATLAB graphical user interface (R. Blaylock, in development) for extracting the articulatory gestures to be analyzed in MVIEW_RT (M. Tiede, in development). Gestures were automatically delimited in MVIEW_RT. For each gesture, the following time points were algorithmically determined from the ROI time functions, using a velocity criterion: onset of movement towards the target, onset of release movement away from target.

The automatic markings generated by MVIEW_RT were compared to the actual movement of the gestures in the MRI video. In case the markings did not match the real articulatory movements, the data was discarded.

The following measures in milliseconds were used in order to test the hypotheses: 1) ConsVons: interval between consonant onset and vowel's oral onset gestures; 2) ConsNons: interval between consonant onset and velum onset gestures; 3) VonsNons: interval between vowel's oral onset and velum onset gestures; 4) CrelNons: interval between consonant release and velum onset; 5) CrelVons: interval between consonant release and vowel's oral onset; 6) VelDur: interval between velum onset and velum release.

ANOVAs were run for each subject for each stimulus set as a function of the word type ($\alpha = 0.5$, cf. table 1). Subsequently, a post-hoc Tukey test was run.

3. RESULTS

Table 2 summarizes the articulatory measures for each syllable type. Table 3 shows that conditions i-v were similar for both subjects when analyzed separately according to the vowel type (cf. table 1). Condition vi was distinct for vowel types "i" and "a". This difference, actually, may be due to prosodic effects. Since the articulatory coordination among the vowel types presented a similar pattern, they were grouped together as seen in table 2.

3.1. ConsVons Coordination

A One-Way ANOVA with ConsVons as function of the word type has shown a statistical difference among the syllables for Speaker 1 ($F(3, 83) = 3.267$, $p < 0.0253$; Speaker 2: n.s.), but a subsequent post-hoc indicated that the only difference was between NV and CVN. Considering that the CV (or CVN) type could have a voiced or a voiceless stop, we suspected this coordination difference might be due to the C voicing type. This observation was partially confirmed by comparing ConsVons as a function of the stop type (Speaker 1: $F(2, 84) = 12.86$, $p < 1.34^{-5}$; Speaker 2: $F(2, 71) = 2.623$, $p < 0.08$). Mean and standard deviation are: i) Speaker 1: [p] = 41 (29); [b] = 8 (27); [m] = 15 (21); ii) Speaker 2: [p] = 59 (30); [b] = 30 (41); [m] = -0.9 (54). Also, according to these results, C and V gestures are roughly synchronous.

3.2. ConsNons Coordination

Either subject presented a statistically different coordination among the syllable conditions (Speaker 1: $F(2,59) = 130.8$, $p < 2^{-16}$; Speaker 2: $F(2,51) = 21.3$, $p < 1.88^{-7}$). All combinations where different except one (NVN = NV), as shown by the post-hoc test. This data also shows that N and V gestures starts roughly at the same time.

3.3. VonsNons Coordination

The coordination for this condition was similar to ConsNons, i.e., CVN ≠ (NVN = NV) (Speaker 1: $F(2, 51) = 108.3$, $p < 2^{-16}$; Speaker 2: $F(2, 51) = 47.75$, $p < 2.06^{-12}$). This data shows that V and N gestures are articulated roughly asynchronous.

3.4. CrelNons Coordination

Although the general pattern here was the same as ConsNons and ConsVons: CVN ≠ (NVN = NV) (Speaker 1: $F(2, 59) = 154.7$, $p < 2^{-16}$; Speaker 2: $F(2,51) = 66.4$, $p < 6.35^{-15}$), the coordination between the consonant release and velum onset seems to be approximately synchronous for CVN. The reason why is because this duration pattern is similar (or quasi similar) to the one in the ConsVons data (Speaker 1: $F(1, 23) = 6.457$, $p < 0.02$; Speaker 2: n.s.), and because the higher values in the ConsVons data are statistically equal to lower values such as 4 (69).

3.5. CrelVons Coordination

The high negative values for this coordination implies that the vowel starts much earlier than the consonant release. All data were similar for Speaker 2 (n.s.) and Speaker 1, with one exception (NV ≠ CVN; $F(3, 83) = 3.649$, $p < 0.0159$). This different is related to the effect discussed in ConsVons.

3.6. VelDur

VelDur was different across syllable types with one exception (NV = CVN; Speaker 1: $F(2, 59) = 26.59$, $p < 5.85^{-9}$; Speaker 2 $(2, 51) = 17.15$, $p < 2.01^{-6}$). This was the only articulatory measure influenced by the prosodic condition: mean and standard deviation for i) **Strong Boundary**: Speaker 1: CVN (161, 37); NV (161, 25); NVN (288, 24); Speaker 2: CVN (164, 34), NV (151, 42), NVN (290, 65); and ii) **Weak Boundary**: Speaker 1: CVN (117, 22), NV (151, 27), NVN (204, 49); Speaker 2: CVN (144, 43), NV (176, 48), NVN (204, 67).

Table 2: Mean and standard deviation of the articulatory measures according to the syllable type. S stands for subject; NA, for not available.

Measure	CV	CVN	NV	NVN	S
ConsVons	22 (28)	**37 (35)**	12 (21)	18 (20)	1
	15 (49)	**34 (48)**	6 (51)	4 (69)	2
ConsNons	NA	145 (43)	-32 (25)	-21 (48)	1
	NA	128 (46)	12 (45)	56 (60)	2
VonsNon	NA	107 (32)	-50 (30)	-39 (45)	1
	NA	146 (54)	-25 (53)	10 (58)	2
CrelNons	NA	**34 (42)**	-147 (30)	-129 (38)	1
	NA	**44 (58)**	-149 (27)	-106 (56)	2
CrelVons	-93 (38)	-75 (31)	-106 (26)	-90 (24)	1
	-117 (48)	-102 (58)	-124 (41)	-117 (60)	2
VelDur	NA	144 (38)	154 (26)	233 (58)	1
	NA	153 (40)	166 (45)	252 (78)	2

Table 3: Statistical significance of the articulatory measures as a function of the word type (CV, CVN, VN, NVN) for each vowel type ([a, i, u]) and subject (1, 2). n.s. = non-significant. * = p < 0.5; ** = p < 0.01; *** = p < 0.001.

Measure	Subject1			Subject 2		
	[a]	[i]	[u]	[a]	[i]	[u]
ConsVons	n.s.	***	n.s.	n.s.	*	n.s.
ConsNons	***	***	***	*	***	**
VonsNons	***	***	***	***	***	***
CrelNons	***	***	***	***	***	***
CrelVons	***	***	***	***	***	***
VelDur	n.s.	***	*	**	***	n.s.

4. DISCUSSION

The data partially support H1 for CVN. Consonant onset and nasal vowel's oral onset are triggered in rough temporal synchrony, i.e., they are considered to be in-phase. Not only the coordination varies around zero (mean NVN: 4 ms; mean CVN: 37 ms. Recall that the coordination for consonant and vowel gestures is similar between these syllables). Nevertheless, we found no evidence for an anti-phase coordination between the oral consonant (labial gesture) and the velum gesture. Despite the fact that the velum movement starts later comparing to the labial movement, it is synced (in-phase) to the consonant release, cf. discussed in sectin 3.4.

Again, visual inspection of the videos gives additional support for this evidence, for the velum gesture starts immediately after the release of the labial gesture. Also, this is not a pattern found in anti-phase consonants, as, for example, words comprising two different syllables (e.g. anti-phase coordination between [p] and [t] in [pa.tʊ]). This same pattern was found for French in Proctor et al.'s study [19, p. 578]: "velum lowering commences soon after the release of the initial labial".

Regarding H2, the results fully support this hypothesis, since either the labial onset-oral onset coordination for CV, the labial onset-oral onset coordination for NV, and the multiple-gesture coordination for NVN are roughly synchronized in time, i.e., in-phase (see table 2).

Finally, as discussed in section 3.1, we need to gather additional data in order to test different CV coordination depending on the C voicing. This effect may be related to the effect of the glottal gesture on the syllable structure, creating another level of complexity to the coordination. Yet, the precise effects of the glottal gesture coordination cannot be directly tested by today's experimental procedures.

5. CONCLUSION

This study provides new insights on the consonant-vowel coordination with nasal vowels in Brazilian Portuguese. The data shows near-synchrony timing (in-phase) for the consonant onset and nasal vowel's oral onset, and the consonant release and velum onset in CVN syllable structure. The other syllables (CV, NV, NVN) present near-synchrony timing (in-phase) for all their gesture's onsets. Moreover, results support the *Coupling hypothesis of Syllable Structure* as an effective and productive way of studying syllables' articulatory coordination.

6. ACKNOWLEDGMENTS

This research was supported by a CAPES grant (10745/13-0) and NIH grants (DC007124, DC03172). The authors would like to thank Asterios Toutios, Sajan Lingala, Colin Vaz, and Yinghua Zhu, for supporting the MRI session, and Christina Hagedorn and Cynthia Lee for helping with the articulatory analysis.

7. REFERENCES

[1] Bresch, E. Nielsen, J., Nayak, K., Narayanan, S. 2006. Synchronized and noise-robust audio recordings during realtime MRI scans, *JASA*, vol. 120, n. 4, 1791–1794

[2] Demolin, D., Delvaux, V., Metens, T., Soquet, A. 2003. Determination of velum opening for French nasal vowels by magnetic resonance imaging. *Journal of Voice.* Dec, 17 (4), 454-467.

[3] Drenska, Margarita. 1988. Análise acústica das vogais nasais em português e búlgaro. In: *Encontro da Associação Portuguesa de Linguística* (APL), 139–165.

[4] Goldstein, L., Byrd, D., Saltzman, E. 2006. The role of vocal tract gestural action in units understanding the evolution of phonology. In: Arbib, M. (ed.), *Action to language via the mirror neuron system*. Cambridge: Cambridge University Press, 215-249.

[5] Goldstein, L., Chitoran, I., Selkirk, E. 2007. Syllable structure as coupled oscillator modes: evidence from Georgian vs. Tashlhiyt Berber. *Proceedings of the XVIth International Congress of Phonetic Sciences* Saarbrucken, Germany, 241-244.

[6] Goldstein L, Nam H, Saltzman E, Chitoran I. 2009. Coupled oscillator planning model of speech timing and syllable structure. In: Fant G, Fujisaki H, Shen J, editors. *Festschrift for Wu Zongji*. Commercial Press; Beijing, 239–250.

[7] Jakobson, R. 1968. *Child Language, Aphasia, and Phonological Universals*. The Hague: Mouton.

[8] Lacerda, A. & Head, B.. 1966. Análise de sons nasais e sons nasalizados do português. *Revista do Laboratório de Fonética Experimental (de Coimbra)* 6, 5–70.

[9] Lacerda, A., Strevens, P. 1956. Some phonetic observations using a speech-stretcher. *Revista do Laboratório de Fonética Experimental (de Coimbra)* 3, 5–16.

[10] Lammert, A., Goldstein, L., Iskarous, K. 2010. Locally-weighted regression for estimating the forward kinematics of a geometric vocal tract model. Proc. Interspeech Makuhari, 1604-1607.

[11] Lammert, A., Ramanarayanan, V., Proctor, M., Narayanan, S. 2013. Vocal tract cross-distance estimation from real-time MRI using region-of interest analysis. *Proceedings of Interspeech 2013*, Lyon, France.

[12] Löfqvist, A., Gracco, V. 1999. Interarticulator programming in VCV sequences: lip and tongue movements. *Journal of the Acoustical Society of America* 105, 1854-1876.

[13] Lovatto, L., Amelot, A., Crevier-Buchman, L., Basset, P., Vaissière, J. 2007. A fiberscopic analysis of nasal vowels in Brazilian Portuguese. *Proceedings of the XVIth International Congress of Phonetic Sciences* Saarbrucken, Germany, 549-552.

[14] Maeda, S. 1993. Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In: Krakow, M. K., Huffman, R. A. (eds). *Nasals, Nasalization and the Velum*. New York: Academic Press, 147-167.

[15] Medeiros, B., Demolin, D. 2006. Vogais nasais do português brasileiro. *Revista da ABRALIN* 5, n.1 and n.2, dec 2006, 131-142.

[16] Montagu, J. 2007. *Analyse Acoustique et Perceptive des voyelles nasales et nasalisées du français parisien*. Thèse de doctorat, Université Paris III - Nouvelle Sourbonne.

[17] Narayanan, S., Nayak, K, Lee, S., Sethy, A., Byrd, D. 2004. An approach to real-time magnetic resonance imaging for speech production, *JASA*, vol. 115, no. 4, 1771–1776.

[18] Oliveira, C. A. M. de. 2009. *Do grafema ao gesto – contributos linguísticos para um sistema de síntese de base articulatória*. Tese (Doutor em Linguística) – Universidade de Aveiro, Aveiro, 2009.

[19] Proctor, M., Goldstein, L., Lammert, A., Byrd, D., Toutios, A., Narayanan, S. 2013. Velic coordination in French Nasals: a Real-time Magnetic Resonance Imaging Study. *Proceedings of Interspeech 2013*, Lyon, France.

[20] Seara, C. 2000. *Estudo acústico-perceptual da nasalidade das vogais do português brasileiro*. Tese de Doutorado, Universidade Federal de Santa Catarina.

[21] Silva, A. R. T. V. 1995. *Análise de Fonemas Nasais da Língua Portuguesa*. Tese de mestrado, Universidade de Aveiro.

[22] Sousa, E. M. de. 1994. *Para a Caracterização Fonético-Acústica da Nasalidade no Português do Brasil*. Dissertação de mestrado em linguística, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Brasil.