AN ARTICULATORY STUDY OF POSTERIOR NASAL DIPHTONGS IN BRAZILIAN PORTUGESE

Rita Demasi[1], Christophe Savariaux[2] & Didier Demolin[1]

Laboratoire de Phonétique et Phonologie, Université de Sorbonne Nouvelle[1] & Gipsa-Lab, Grenoble[2]
rita-de-cassia.demasi@etud.sorbonne-nouvelle.fr, christophe.savariaux@gipsa-lab.grenoble- inp.fr,
didier.demolin@univ-paris3.fr,

ABSTRACT

Nasal diphthongs are quite rare in the world's languages. This paper analyzes how speakers control articulatory movements for nasal diphthongs in Brazilian Portuguese (BP). Our aim is to characterize the oral-nasal coupling in posterior nasal diphthongs from the Paulistano dialect spoken in the city of Sao Paulo. We show that oral and nasal diphthongs have different tongue contours, besides velopharyngeal coupling. A 2D EMA study was carried out to contrast [aw] and [ãw̃] in monosyllabic words.

Keywords: Brazilian Portuguese, Nasal Diphthongs, EMA, Articulatory Targets, Tongue, Velum.

1. INTRODUCTION

Nasals in diphthongs are rare phonemes in the world's languages. These sounds are analyzed as vowel segments combined with a velum gesture that have specific acoustic features due to the oro-nasal coupling. However, the binary feature [± nasal] doesn't take into account the interactions between articulatory movements and the vocal tract geometry in speech production. An analysis in terms of binary features doesn't consider the effects of coarticulation between the trajectory of the tongue and velum opening-closing phases. The velopharyngeal port activity varies across languages [1] both in anticipatory or carryover coarticulation. In this study we assume that *time* is a relevant dimension and that it is a primitive variable of the phonological representations [2]. In Brazilian Portuguese the oral/nasal opposition is distinctive in back diphthongs. Thus, we compared the speech production of [ãw̃] vs [aw]. Our goal is to analyze nasal diphthongs from the Paulistano dialect. These sounds will be considered as dynamic articulations influenced by gestural coarticulation effects and their possible overlap. To provide dictions between oral and nasal diphthongs at the articulatory level we made a 2D EMA experiment. The goal is to track the spatial-temporal relation between the velopharyngeal port and the tongue. Our hypothesis is that the oral gestures in posterior nasal diphthongs have a different timing when oral and nasal diphthongs are compared. If tongue movements are different in nasal and oral diphthongs, the trajectories of gestures are not realized in the same way. The main goal of this study is therefore to characterize production differences between oral and nasal diphthongs. What kind of temporal and spatial parameters must speakers control to minimize the articulatory effort and to maximize perception [3]? This paper is organized with one section showing the recording methodology. The rest of the paper presents results, discussion and preliminary conclusions of this ongoing research.

2. METHOD

2.1. Speakers

Two female speakers participated in this experiment (S1 e S2). Both are thirty years old and they are native speakers of the São Paulo city dialect of Brazilian Portuguese. Both speakers are living in France since less than two years.

2.2. Material

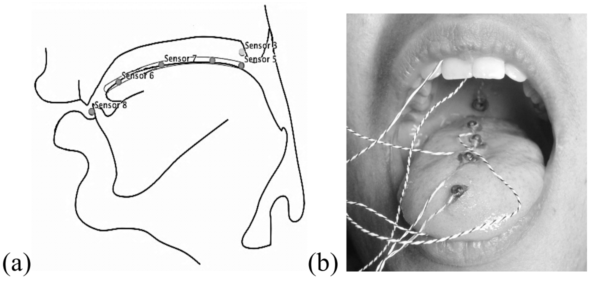
Ten monosyllabic words from Brazilian Portuguese were recorded. All have a CVG and a CṼG̃ sequence which includes the diphthongs. The *corpus* contrasts oral and nasal diphthongs in minimal pairs, e.g. [paw] vs [pãw̃]; [saw] vs [sãw̃]; [maw] vs [mãw̃]; [taw] vs [tãw̃] and [kaw] vs [kãw̃]. All words were recorded in a carry sentence: [dʒigʊ__todʊ dʒiɐ]. We set a voiceless dental consonant /t/ after the diphthong to minimize the articulatory boundary effects at the end of the diphthongs.

2.3. Data processing

Data were collected at the Gipsa_Lab of Grenoble-Alpes University, in France. To map the intraoral articulator movements, data were recorded using a Midsagittal Electromagnetic Articulograph (Carstens AG100). Thus, the articulatory movements were recorded using pellets attached on the following parts of the tongue: tongue tip (TT) (a), tongue blade

(TBL) (b), tongue dorsum (TD) (c), tongue back (TBCK) (d) and velum (VEL) (e), plus the nose (NS) (f), the upper incisors (UI) and lower incisors (JAW) (g). The pellets are placed along the mid-sagittal line and they are spaced at a distance of approximately 1,5 cm. The pellet on the velum is at a distance of approximately 1,2 cm of the back part of the tongue dorsum as shown in Figure 1.

Figure 1: (a) Medio-sagittal cut showing the 7 pellets position. (b) Photography of the Ema pellets position on the tongue of one of the speakers.



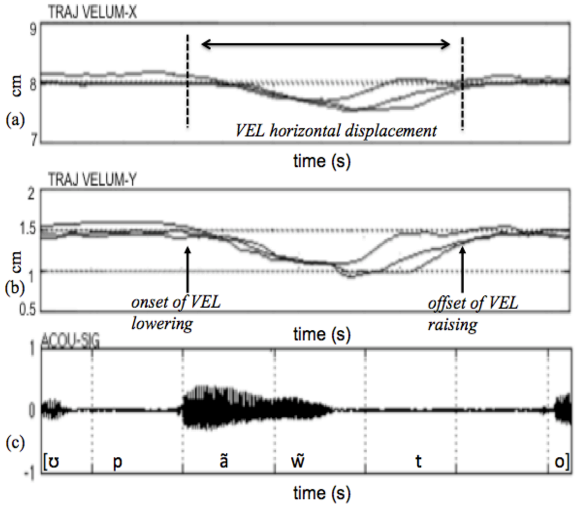
(a) (b)

EMA data were synthetized by Chebyshev low-pass filtered on Matlab. The head movement and rotation variation are determined by the helmet position. The line determining the boundary of the soft and hard palate was made by running a pellet from the back of the soft to the hard palate. The occlusal plane (bite plane) oriented the zero bi-dimensional position in *x-axis* (horizontal) and *y-axis* (vertical). Acoustic stimuli and articulatory data were collected synchronically in the same temporal scale. In the recording session the sentences were presented randomly on a personal computer screen and each participant repeated 5 times the stimuli and we asked to the subject speak in a normal speech rate. Each subject produced 50 tokens. EMA data were processed with Matlab through homemade software TRAP [4]. A *t*-test was used to compare the intra-speaker variation in the set of the data.

3. RESULTS

Our measurements were made from the variation of articulatory trajectories of oral and nasal diphthongs. All the measures have a reference point, the zero crossing point in *x-y*. To infer the oral tract spatial variation the delta between the occlusal plane and the higher point of the hard palate was calculated. On the horizontal orientation, we calculated the delta between the UI and the end point of the soft palate. Values for S1 is: 6,5 cm on the *x*-axis and 2,10 cm on the *y*-axis, for S2 *x*-axis = 7,0 cm and *y*-axis = 2,40 cm. Figure 2 shows the visual presentation of the VEL gesture trajectories. This also illustrates the criteria for the extraction of quantitative parameters. The arrows identify the onset/offset of the velum gesture.

Figure 2: 3 examples of the nasal diphthong [ɐ̃w̃] trajectories of speaker S1. (a) VEL trajectories displacement are superimposed and aligned in the same window time on the *x-axis*. (b) VEL trajectories displacement are superimposed and aligned in the same window time on the *y-axis*. (c) A waveform from one acoustic stimulus.



The onset represents the VEL pellet stationary position, before the start of the velum lowering to produce the nasal diphthongs. The lower position characterizes the largest amplitude displacement of the VEL pellet. The offset is the stationary position after the velum-raising phase. Table 1 and 2 shows the raw data that displays the distances of tracking points for the VEL pellet movement.

Table 1: Mean of the delta of the static spatial position (mm) for velum tracking displacement from the stimuli [ɐ̃w̃] on the *x* and *y* axes to *n* = 25. STEDV in parentheses.

Word		VEL pellet spatial variation (mm)			
		TRAJ VEL X		TRAJ VEL Y	
S1	[pɐ̃w̃]	25.6	(7.6)	32.7	(10)
	[mɐ̃w̃]	28.6	(7.6)	27.6	(13.4)
	[kɐ̃w̃]	22.9	(6.7)	27.4	(10.4)
	[sɐ̃w̃]	26.6	(10.1)	33.7	(9.6)
	[tɐ̃w̃]	26.6	(8.9)	33.7	(8.8)
	mean	26.4	(8.5)	30.7	(10.9)
S2	[pɐ̃w̃]	20.2	(1.8)	25.2	(3.1)
	[mɐ̃w̃]	19.3	(5.6)	26.4	(5.6)
	[kɐ̃w̃]	19.3	(5.1)	26.4	(8.3)
	[sɐ̃w̃]	19.4	(2.7)	27.5	(5.0)
	[tɐ̃w̃]	21.1	(1.6)	31.2	(2.2)
	mean	19.4	(4.0)	25.7	(6.7)

Table 2: Mean distances of the static position of the trajectories (cm) of velum onset, lower position and offset from the stimuli [ɐ̃w̃] on the *x* and *y* axes for *n* = 25. Speaker S1. STEDV in parentheses.

	Word	VEL X		VEL Y	
Onset	pão	8.21	(0.07)	1.65	(0.10)
	mão	8.20	(0.05)	1.63	(0.05)
	são	8.19	(0.03)	1.59	(0.07)
	tão	8.18	(0.02)	1.60	(0.04)
	cão	8.19	(0.02)	1.59	(0.01)
	mean	8.19	(0.01)	1.61	(0.02)
Lower	pão	7.96	(0.05)	1.33	(0.03)
	mão	7.91	(0.07)	1.35	(0.13)
	são	7.96	(0.07)	1.32	(0.12)
	tão	7.91	(0.09)	1.27	(0.09
	cão	7.96	(0.10)	1.32	(0.08)
	mean	7.94	(0.02)	1.32	(0.03)
Offset	pão	8.18	(0.02)	1.58	(0.04)
	mão	8.18	(0.08)	1.59	(0.11)
	são	8.22	(0.04)	1.64	(0.07)
	tão	8.21	(0.01)	1.63	(0.02)
	cão	8.22	0.03)	1.64	(0.06)
	mean	8.20	(0.02)	1.62	(0.03)

Table 3: Mean of the static position of the trajectories (cm) of velum onset, lower position and offset from the stimuli [ɐ̃w̃] on the *x* and *y* axes for *n* = 25. Speaker S2. STEDV in parentheses.

	Word	VEL X		VEL Y	
Onset	pão	5.09	(0.01)	1.03	(0.03)
	mão	5.08	(0.07)	0.97	(0.07)
	são	5.07	(0.02)	1.03	(0.04)
	tão	5.09	(0.02)	1.06	(0.04)
	cão	5.08	(0.06)	1.01	(0.09)
	mean	5.08	(0.01)	1.02	(0.03)
Lower	pão	4.89	(0.02)	0.78	(0.01)
	mão	4.91	(0.04)	0.78	(0.04)
	são	4.88	(0.01)	0.75	(0.01)
	tão	4.88	(0.01)	0.75	(0.01)
	cão	4.88	(0.01)	0.75	(0.01)
	mean	4.89	(0.01)	0.76	(0.02)
Offset	pão	5.07	(0.02)	1.01	(0.03)
	mão	5.08	(0.02)	1.01	(0.05)
	são	5.06	(0.03)	0.99	(0.04)
	tão	5.07	(0.03)	1.01	(0.04)
	cão	5.04	(0.04)	0.96	(0.07)
	mean	5.06	(0.01)	1.00	(0.02)

The temporal values in Table 4 are the delta between velum onset and velum lower position, and the delta between velum lower position and velum offset.

Table 4: Mean of duration between pellets (ms) position for velum tracking displacement from the stimuli [ɐ̃w̃] on the *x* and *y* axes to *n* = 25 STEDV in parentheses in parentheses.

	Word	Velum pellet temporal duration (mm)			
		S1		S2	
Lowering	pão	186	(29)	228	(41)
	mão	208	(34)	328	(44)
	são	163	(64)	215	(55)
	tão	225	(37)	276	(44)
	cão	163	(76)	268	(25)
	mean	204	(58)	263	(59)
Raising	pão	157	(56)	124	(21)
	mão	192	(81)	177	(92)
	são	204	(56)	165	(30)
	tão	192	(31)	155	(22)
	cão	204	(34)	169	(29)
	mean	191	(58)	158	(51)
Total displacement	pão	343	(73)	352	(29)
	mão	400	(65)	505	(82)
	são	367	(90)	380	(79)
	tão	417	(41)	431	(42)
	cão	367	(68)	437	(41)
	mean	395	(79)	421	(79)

Figure 3: 1 repetition of the oral diphthong stimulus [aw] in the word *pau* for speaker S1. (a) VEL trajectory on the *y-axis*. (b) TBCK trajectory on the *x-axis*. (c) TBCK trajectory on the *y-axis*. (d) The spectrogram is from the acoustic recording.

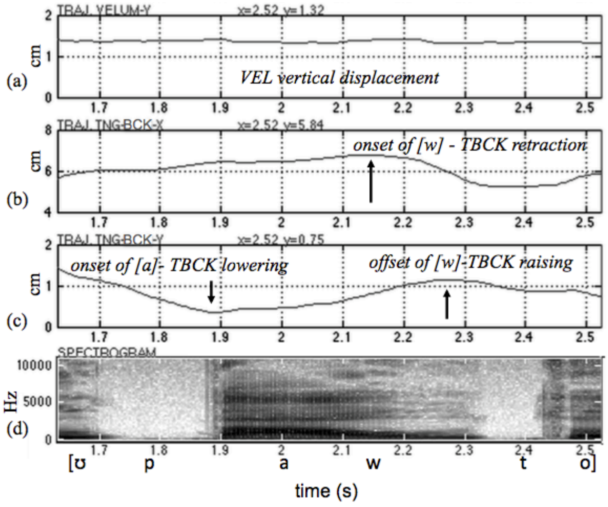


Figure 3 identifies the onset/offset of the tongue gesture movement in oral diphthongs. Figure 4 and 5 shows the pellets movement in oral cavity during oral and nasal diphthong production.

Figure 4: 1 repetition of the oral diphthong [aw] in the word *cau* from speaker S2. Pellets trajectories on the *x-y* axes (cm).

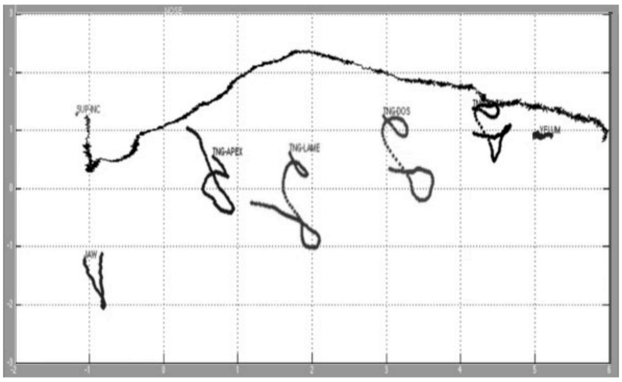
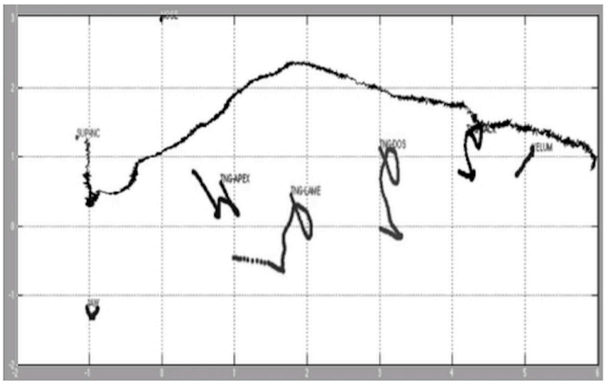


Figure 5: 1 repetition of the oral diphthong stimulus [ãw̃] in the word *cão* from speaker S2. Pellets trajectoie on the *x-y* axes (cm).



4. DISCUSSION

The analysis of our data suggests that nasal diphthongs are produced in two phases and that the tongue has an important dynamic role. The gestures of the diphthongs have two targets with different height and constriction [5]. Even though the mean values of velum movement duration seems different for S1 and S2, a *test-t student* shows: two velum lowering movements $p = .9$; two velum raising displacement $p = .3$ and the total duration of velum activity $p = .9$. So, the velum duration must be one important parameter for the perception of nasality. We detected differences in the degree of velum lowering. For S1 and S2, $p=.004$ and $p=.06$ on x-y axes. To produce a posterior nasal diphthong, besides velum activity, we observed that the tongue transition between targets is more rapid and continuous than in nasal than in oral diphthongs (cf. Fig.4 and Fig. 5). The velum movement is a parameter implying a change in the geometry of the vocal tract. However the velum movement is not the only gesture responsible to produce the oral nasal distinction. As shown by Cagliari [6] all nasal vowels of BP are diphthongized. So, when the velopharyngeal aperture is active, the oral constriction seems to contribute to a drop in nasal airflow [7] and this changes the spectral properties of the sound. Maybe, the control of the timing between the tongue and the velum gestures is more important to judge the perception of the nasal quality in diphthongs than only the oro-nasal coupling [8].

5. REFERENCES

[1] Bell-Berti, F. 1993. Understanding velic motor control: studies of segmental context. In: Huffman, M.K, Krakow, R. A. (eds), *Phonetics and Phonology*, Volume 5: Nasals, nasalization, and the velum, 63-85.

[2] Fowler, C.A., Rubin, P.,Remez, R.E.and Turvey, M.T. 1980. Implications for Speech Production of a General Theory of Action. In *Language Production, Volume 1 Speech and Talk*. B. Butterworth (Ed.) New-York: Academic Press.

[3] Demolin, D. 2002. The search for primitives in phonology and the explanation of sound patterns: the contribution of fieldwork studies, In: Gussenhoven, C. Warner, N. (eds), *Papers in Laboratory Phonology 7*. Mouton de Gruyter, Berlim. 455-513.

[4] Savariaux, C. 2014. *Trap: Logiciel de traitement des signaux de parole*. [Manual]. Version 6. Département Parole et Cognition, GIPSA-lab, Grenoble-Alpes University – France.

[5] Stevens, K. 1999. Articulatory-acoustic-auditory relationships. In: Hardcastle, W., Laver, J. (eds), *The Handbook of Phonetic Science*. Oxford: Blackwell, 462-506.

[6] Cagliari, L.C. 1977. *An experimental study of nasality with particular reference to Brazilian Portuguese*. Edimburgo, PhD Dissertation. Linguistic Department of Edinburg University.

[7] Demasi, R.C.B. 2010. *A ditongação nasal no português: uma análise acústico-aerodinâmica da fala*. Master Dissertation. São Paulo University. USP.

[8] Kingston, J., Diehl, R. L. 1994. Phonetic knowledge. *Language* 70, 419-454.