European Portuguese MRI Based Speech Production Studies

Paula Martins, Inês Carbone, Alda Pinto, Augusto Silva, António Teixeira

To cite this version:

Paula Martins, Inês Carbone, Alda Pinto, Augusto Silva, António Teixeira. European Portuguese MRI Based Speech Production Studies. Speech Communication, 2008, 50 (11-12), pp.925. 10.1016/j.specom.2008.05.019 . hal-00499224

HAL Id: hal-00499224

https://hal.science/hal-00499224

Submitted on 9 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

European Portuguese MRI Based Speech Production Studies

Paula Martins, Inês Carbone, Alda Pinto, Augusto Silva, António Teixeira

PII: S0167-6393(08)00080-0
DOI: 10.1016/j.specom.2008.05.019
Reference: SPECOM 1722

To appear in: *Speech Communication*

Received Date: 10 June 2007
Revised Date: 11 April 2008
Accepted Date: 21 May 2008

Please cite this article as: Martins, P., Carbone, I., Pinto, A., Silva, A., Teixeira, A., European Portuguese MRI Based Speech Production Studies, *Speech Communication* (2008), doi: 10.1016/j.specom.2008.05.019

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

European Portuguese MRI Based Speech Production Studies ⋆

Paula Martins [a] Inês Carbone [b] Alda Pinto [c] Augusto Silva [b] António Teixeira [b,*]

[a] *Escola Superior de Saúde, Universidade de Aveiro, Portugal*

[b] *Dep. Electrónica Telec. Informática/IEETA, Universidade de Aveiro, Portugal*

[c] *Dep. de Radiologia, Hospital da Universidade de Coimbra, Portugal*

Abstract

Knowledge of the speech production mechanism is essential for the development of speech production models and theories. Magnetic Resonance Imaging delivers high quality images of soft tissues, has multiplanar capacity and allows for the visualization of the entire vocal tract. To our knowledge, there are no complete and systematic Magnetic Resonance Imaging studies of European Portuguese production. In this study, a recently acquired Magnetic Resonance Imaging database including almost all classes of European Portuguese sounds, excluding taps and trills, is presented and analyzed. Our work contemplated not only image acquisition but also the utilization of image processing techniques to allow the exploration of the entire database in a reasonable time. Contours extracted from 2D images, articulatory measures (2D) and area functions are explored and represent valuable information for articulatory synthesis and articulatory phonetics descriptions. Some European Portuguese distinctive characteristics, such as nasality are addressed in more detail. Results relative to oral vowels, nasal vowels and a comparison between both classes are presented. The more detailed information on tract configuration supports results obtained with other techniques, such as EMMA, and allows the comparison of European Portuguese and French nasal vowels articulation, with differences detected at pharyngeal cavity level and velum port opening quotient. A detailed characterization of the central vowels, particularly the [ɨ], is presented and compared with classical descriptions. Results for consonants point to the existence of a single positional dark allophone for [l], a more palatoalveolar place of articulation for [ʎ], a more anterior place of articulation for [ʎ] relative to [ɲ], and the use, by our speaker, of a palatal place of articulation for [k]. Some preliminary results concerning coarticulation are also reported. European Portuguese stops revealed less resistant to coarticulatory effects than fricatives. Among all the sounds studied, [ʃ] and [ʒ] present the highest resistance to coarticulation. These results follow the main key features found in other studies performed for different languages.

Key words: Speech Production, European Portuguese, Magnetic Resonance

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Imaging, Nasals, Coarticulation

1 Introduction

Mankind's knowledge about Human speech production and perception is still incomplete. More information is definitely needed. Recently, better techniques for measuring vocal tract configurations have become an increased research interest. Building phonetic information databases has had great relevance in fields such as speech synthesis, speech recognition, speech disorder studies, learning of new languages, etc. An area where production data are very important is articulatory synthesis, where we have been involved for more than a decade [1]. These anthropomorphic synthesizers demand large amounts of detailed anatomic-physiological information, if possible in 3D, and their variation in time (dynamic information). For European Portuguese (EP), not much information is available.

To compensate this lack of information, the objectives of the present study are: 1) to provide vocal tract configurations during (sustained) production of all the EP sounds (excluding taps and trills); 2) to perform comparisons between different sound classes; 3) to obtain direct area functions from a great part of the EP sounds; 4) to have a preliminary approach on coarticulation in stops and fricatives and, 5) due to the nature of the research team, to develop acquisition and segmentation techniques with application in the field of speech production.

This paper is structured as follows: this first section introduces the problem, presents the most common anatomic-physiological measurement methods for speech production studies, describes the EP relevant specificities, coarticulation and related work in MRI application to speech production studies; section 2 describes image acquisition and corpus; section 3 describes image processing; sections 4 and 5 present our results, separated into vowels and consonants. All the phonetic considerations made in this paper rely on static articulations that might be different from continuous speech articulations. The paper ends with a discussion of the results presented in earlier sections, and with the main conclusions that can be extracted from them.

⋆ Part of the work reported, particularly on nasals, was accepted for presentation at Interspeech 2007. Paper is entitled "An MRI study of European Portuguese nasals".
∗ Dep. Electrónica Telec. Informática/IEETA, Universidade de Aveiro, 3810 AVEIRO, PORTUGAL
Tel: +351234370500 Fax: +351234370545 Email: ajst@ua.pt

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Nomenclature:

ETL	Echo Train Length	FLASH	Fast Low Angle Shot
FOV	Field of View	MPRAGE	Magnetization Prepared Rapid Acquisition Gradient Echo
MRI	Magnetic Resonance Imaging	NEX	Number of Excitations
SSFP	Steady State Free Precession	TE	Time to Echo
TR	Time to Repeat	TSE	Turbo Spin Echo (sequence)
VIBE	Volume Interpolated Breath Hold Examination	VPOQ	Velum Port Opening Quotient

1.1 Measurement methods

Nowadays, the common methods found in the speech research literature to acquire anatomic-physiological information directly are: Electromagnetic Midsagittal Articulography (EMMA), Electropalatography (EPG), and Magnetic Resonance Imaging (MRI). EMMA provides valuable kinematic data relative to different articulators (lips, tongue, jaw, velum) with good temporal resolution. However, some drawbacks can be pointed out: the acquired data are, in the majority of available systems, two dimensional and limited to the trajectories of some articulator fleshpoints [2,3]; the process is invasive and articulation may be affected by the sensors. EPG measures only the linguopalatal contact and its variation on time, being difficult to make well-fitted pseudo-palates, which in turn interfere to some extent with speech production [4].

MRI, the technique on which we will focus in this study, has some potential advantages: it provides a good contrast between soft tissues, allows 3D modeling and covers the vocal tract in all of its extension [5,6,7,8]. This last advantage is of special interest in the study of the pharyngeal cavity, as it is not accessible through EMMA or EPG. Moreover, it is non-invasive and considered as safe. Its disadvantages are related to the absence of the teeth in the images, due to their lack of Hydrogen protons; the acquisition technique, in which the speaker must be lying down during speech production. This position can have some influence, for instance, on the tongue posture [9,10], but this drawback can be considered acceptable.

The relatively low temporal resolution achieved, even with the fastest acquisition techniques, is a limiting factor [8]. The noisy acquisition environment and the reduced acoustic feedback, due to the use of headphones, are also MRI disadvantages.

The MRI technique has already been used for the study of several languages:

ACCEPTED MANUSCRIPT

British English [5], American English [11,12,7,8], French [13,14,15], Swedish [16,17], Japanese [18], German [19,20,21], Tamil [8], and Akan [22]. For EP, one of the authors was involved in the creation of the first and, to the best of our knowledge, unique EMMA database focused on nasals [23]. Also, there are no EPG databases for EP, and there is only one partial MRI study, [24]. For Brazilian Portuguese this information is also scarce. An MRI based study of nasals was performed recently by Gregio [25].

1.2 European Portuguese

"The characteristics which at first hearing distinguish the pronunciation of Portuguese from that of the other Western Romance languages [are]: (a) the very large number of diphthongs (...); (b) the large number of nasal vowels and nasal diphthongs; (c) frequent alveolar and palatal fricatives (...); (d) the extremely 'dark' quality of the common variety of l-sound" [26, p. 6]. Despite its similarities to Spanish, both in vocabulary and grammatical structure, Portuguese differs considerably in its pronunciation [26].

In EP there is a maximum of 9 oral vowels and ten oral diphthongs [27]. Oral vowels are usually divided into: anterior ([i], [e], and [ɛ]); central ([a], [ɐ], and [ɨ]); and posterior ([u], [o], and [ɔ]). The most problematic vowel is [ɨ] with descriptions going from the schwa to a high central vowel or even, as proposed by Cruz-Ferreira [27], a configuration close to [u]. EP has five nasal vowels ([ĩ], [ẽ], [ɐ̃], [õ], and [ũ]); three nasal consonants ([m], [n], and [ɲ]); and several nasal diphthongs and triphthongs. Despite nasality being present in most of the languages of the world, only about 20% of such languages have nasal vowels [28]. There is some uncertainty in the actual configurations assumed by the tongue and other articulators during EP nasals production, namely nasal vowels. This is particularly relevant for mid vowels where the opposition between mid-low and mid-high, present in the oral vowels set, is neutralized [29]. This neutralization allows the oral articulators to rearrange, leading to associate each nasal vowel to several possible oral counterparts [29]: nasal vowel [ẽ] relates to [e] and [ɛ]; [õ] relates to [o] and [ɔ]; and [ɐ̃] can be more open than [ɐ] or produced with an oral configuration similar to [a]. Note that [i] and [u] are considered to be the oral counterparts of [ĩ] and [ũ]. Also, some phonetic studies point to the existence of differences related with production of EP nasals relative to French [30,23]. In this work, we return to the same challenging topic, using MRI as the data acquisition method.

In EP six fricative consonants are described [31]. Three are produced with vocal fold vibration (voiced fricatives [v], [z] and [ʒ]) and three produced without vibration (unvoiced fricatives [f], [s] and [ʃ]). Sounds [v] and [f] are produced with a constriction point induced by the contact of the lower lip and upper

ACCEPTED MANUSCRIPT

incisor (labiodental), [s] and [z] are fricatives produced with approximation of the tongue tip or blade to the alveolar region. Finally, [ʃ] and [ʒ] are produced in the palato-alveolar area. Phonologically EP has two laterals, /l/ and /ʎ/. The former is produced with contact of tongue tip or blade in the alveolar ridge, the latter produced with a central occlusion between the most anterior tongue dorsum and the anterior palate (palatal consonant).

For the apical lateral /l/, in accordance with EP most frequent descriptions, two allophones are considered: one, non-velarized light or clear [l], occurring in syllable onset; the second, occuring in coda or in absolute word-final position, considered a "velarized" [ɫ] and corresponding to the descriptions of the English dark [l]. During the production of this dark allophone, a second and posterior constriction, originated by tongue back raising towards the velum, is considered [32]. However, Andrade [33] found in three Lisbon speakers, evidence that this "velarization" can also occur in syllable onset. This was also described, much earlier, in older EP phonetic descriptions (Strevens, 1954). Also, Recasens and Espinosa (2005) [34], based on acoustic data stated that EP, together with Russian and Leeds British English, belong to a group of sound systems where /l/ presents the same realization in word initially and word finally.

1.3 Coarticulation

The term coarticulation has been introduced by Menzerath and Lacerda - a Portuguese Phoneticist - in 1933 [35]. Although it could be simply defined as "the articulatory or acoustic influence of one segment or phone on another" [36] it is a complex and difficult subject. Many theories and models have emerged to explain coarticulation but some doubts still persist. There are, however, some accepted facts: coarticulation was observed in almost all languages, being a universal phenomenon, but coarticulatory effects vary from one language to another [37, p.180]. Recent theories of speech production consider that coarticulation plays a central role and that is essential to take coarticulatory effects into account in both speech production models and speech synthesis. Important concepts such as "coarticulation resistance" and "degree of articulatory constraint" (DAC) were introduced to explain why coarticulatory effects are different in different sounds [38]. To give a complete picture of coarticulation one should consider lingual, jaw, labial, and laryngeal coarticulation. An extensive review of the subject can be found in [39].

Several exploratory techniques are referred as important tools when studying coarticulation, such as EMMA [2,3] or EPG [40]. MRI has also been used for the same purpose as described in [12,41,42] and [10]. We are not aware of any MRI coarticulation study for EP.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

1.4 MRI in speech production studies: an overview

MRI evaluation of the vocal tract configuration is definitely not a recent issue in the field of speech production. One of the pioneer studies in this field was performed by Baer et al. [5] for British English. Although it is not the first study that employs MRI as an imaging tool, it was the first that allowed extraction of valuable 3D information related with English vocalic sounds [43].

Traditionally, studies involving MRI were called static (2D and 3D), or dynamic/real-time, although different terminology has been used by different authors, as has been pointed out and explained by Narayanan et al. [8]. From static (2D and 3D) studies, with images acquired during sustained production of sounds, midsagittal profiles and distances, cross sectional areas, articulatory measures, vocal tract area functions, and 3D visualizations were obtained [5,44,16]. The acquisition time, during which articulation must be sustained, is nowadays substantially shorter in most recent studies, when compared with the first MRI evaluations, which reflect technical advances in the field of MRI technology. This fact leads to a better image quality, since image artifacts, due to movements, contributes negatively to the sharpness and image contrast in a MRI image. For real-time studies, recent improvements in temporal resolution are encouraging, but not yet enough to obtain dynamic information relative to some articulators (e.g. tongue tip or velum opening/closure during nasals sounds), or to study more demanding sounds in terms of temporal resolution as happens with stops [21].

The number of speakers participating in studies with published results is not high, varying between one [45,44,46,16,47,19,48], two [5,22,17], four [49,7,6,50,51] and five [52]. This fact reflects the high costs of MRI equipment and the access constraints imposed by the use, in the majority of the studies, of hospital diagnostic equipment. There are studies for different languages and for different classes of sounds. In the next paragraphs, one for each class of sounds contemplated in the present study, a brief review of studies, having a phonetical speech production point of view, is made.

Oral vowels were studied for American English, [44], British English, [5], Akan, [22], Japanese, [53], French, [50], German, [20] and Swedish, [54,17]. Common results are MRI images, distances, segmentations, 3D vocal tract and tongue visualizations, and area functions.

Nasal vowels were mainly considered for French [55,51,56]. In Demolin et al. [55] the results presented are transversal MRI images, cross sectional areas, comparisons between oral and nasal vowels, and 3D reconstructions of the pharynx and of the nasal tract. In 2002, Delvaux et al. [57], obtained from

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

MRI images the articulatory contours. Recently, Engwall et al. [56] published MRI images, nasal and oral areas and a relative measure for the velum port opening, VPOQ.

Dang and his colleagues [58,52] studied nasal consonants for Japanese, Story et al. [44] for (American) English, and Hoole et al. [20] for German. Japanese studies presented several measurements of the three-dimensional geometry of the vocal tract. In [44] area functions and vocal tract visualizations are presented. Hoole and coworkers provided tongue contours and respective deformations based on a two-factor tongue model.

The study lead by Story [44], included some investigation on American English stops, through the observation of 3D vocal tract visualizations and respective area functions. Hoole et al. [20], in 2000, acquired MRI coronal, axial and sagittal volumes of long German vowels and alveolar consonants. Kim [59] studied Korean coronal stops and affricates. She presented midsagittal MRI images, tongue contours, and some measurements of movements, distances, and widths.

Fricatives were studied for a broad number of languages, such as English (British and American), Swedish, German. The oldest study, by Shadle et al. [60] in 1996, showed only midsagittal MRI images. Mohammad et al. [61] developed a new method to acquire MRI dynamic images. Jackson [62], in his work on acoustic modeling, used MRI to draw contours and area functions. Narayanan and Alwan [63] used vocal tract area functions obtained from MRI images of voiced and unvoiced English fricatives to delineate hybrid source models for fricative consonants. Engwall and Badin [41] presented midsagittal contours, 3D vocal tract shapes and investigated coarticulatory effects in Swedish fricatives. Hoole and his team [20] focused on the study of the tongue.

To gather data on laterals, and to the best of our knowledge, Bangayan et al. [64], Narayanan et al. [7], Gick et al. [65] (for American English) and Hoole et al. [20] (for German) used MRI. They presented coronal MRI images, midsagittal segmentations of the vocal tract, area functions, 3D vocal tract and tongue visualizations.

2 Image acquisition

2.1 MRI acquisition

The MRI images were acquired using a 1.5 Tesla (Magneton Simphony, Maestro Class, Siemens, Erlangen, Germany) scanner equipped with Quantum

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

gradients (maximum amplitude=30 mT/m; rise time=240 μs; slew rate=125 T/m/s; FOV=50 cm). Neck and brain phased array coils were used.

Two different types of acquisitions were performed, 2D static and 3D static, whose acquisition sequence parameters are shown in Table 1.

For 3D, instead of exciting a series of 2D slices in different planes (coronal, coronal oblique and axial) as reported by other authors in the field (e.g [14,16]) we performed a volumetric acquisition, by exciting a volume of spins in the axial plane (from above hard palate level to C5-C6 level), using a three-dimensional Fourier Transform (3DFT) sequence. This acquisition has some advantages when compared with 2D acquisitions: the possibility of having a reduced slice thickness (in our study we obtained an effective slice thickness of 2 mm) contributing to obtain high resolution images with a reduced acquisition time; Signal to Noise Ratio (SNR) is usually high with a 3D excitation; possibility of reslicing in any direction with different slice thickness, a variable number of slices and different orientation with a quality superior that can be obtained with 2D acquisitions. When 3D visualizations are required, this method allows the utilization of faster and direct segmentation tools (e.g itk-SNAP) to extract tract configuration. Establishing some trade-offs, we obtained at least the same amount of data as reported in the referenced studies, with a reasonable spatial resolution, but decreasing to less than half the acquisition time (18s).

Bidimensional acquisitions resulted in images of 256x256 pixels and a resolution of 0.78 mm/pixel in both directions. For 3D, the volume has 512x416x60 voxels and resolution of 0.53 mm/pixel in plane and 2 mm resolution in the z direction.

2.2 Corpus

The corpus comprises two subsets, 2D and 3D corpus, acquired using two different acquisition techniques. In both sets, the sounds are artificially sustained (vowels) or holding the articulation (stops) during the period of image acquisition, as already done in a similar way for other languages [44,13,16]. Although with some technical differences, our 2D and 3D corpus were inspired by the studies of [50] for French, [14] also for French, and [16] for Swedish. As in [16], we decided to obtain a large corpus with only one speaker rather than to obtain a small set of items relative to vowels or classes of consonantal sounds with a higher number of speakers. The reason for this option relies on the scarcity of MRI information for EP. Both approaches present advantages and limitations as emphasized by Engwall and Badin [16].

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1
MRI sequence parameters used in imaging acquisition.

Parameter	TSE T1 weighted (2D)	3D flash VIBE
TR (Time to Repeat)	400 ms	4.89 ms
TE (Time to Echo)	8.3 ms	2.44 ms
ETL	15	1
FA	180^{o}	10^{o}
FOV (x,y) [mm]	200 x 200	270 x 216
Slabs	-	1
Slices per slab	-	60
Slice thickness	5 mm	2 mm
Orientation	Sagittal	Axial
Distance factor	-	0.2 mm
Base resolution	256 mm	256 mm
Phase resolution	75%	60%
Phase direction	Ant.-Post.	Right-Left
Phase partial Fourier	-	6/8
BW (Hz/pixel)	235	350
Acquisition time	5.6 s	18 s
NEX	1	1
Image size (x,y) [pixels]	256 x 256	512 x 416
Pixel size (x,y) [mm]	0.78 x 0.78	0.53 x 0.53
Number of measurements	1	1

2D corpus: The main goals were: to obtain MRI static images of the vocal tract during the production of all EP vowels and consonants allowing to extract midsagittal contours; to have articulatory measures; and to measure midsagittal distances. Each sound of the 2D corpus (Table 2) was pronounced and sustained during the acquisition time (5.6 s). To help the speaker, a reference word, containing the target phone, was presented before launching the sequence, using the intercom (e.g "please say [a] as pronounced on [patu]"). This procedure was used for oral and nasal vowels, nasals, laterals and fricatives with one sample of each sound. For nasal vowels this process does not take into consideration the reported dynamic movement between an oral position towards a nasal position (see for example [23,30]). The acquired image should be considered as more representative of nasal vowels when produced in isolation and of the initial and medial configuration during nasal vowel

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

production. To allow a coarticulation study, stops and fricatives were also acquired on a Vowel-Consonant-Vowel (VCV) symmetric context (non-sense words), with V being one of the cardinal vowels [a, i, u]. Note however that, due to recording duration constraints and the secondary role of coarticulation study in the present paper, only stops and fricatives were considered here.

During this recording sequence the speaker was instructed to perform the VC-transition, then to sustain the consonant during acquisition time, and finally perform CV transition. Acquisition was started as soon as the speaker started producing the consonant; the speaker used the acquisition noise to make the final transition. The speaker had the opportunity of having a small training phase before the image acquisition session.

3D Corpus: For this corpus the main purposes were: (1) to obtain tridimensional information, such as vocal tract area functions, and (2) to complement the 2D information with lateral information.

The main challenge with this corpus was to obtain a large volume of data within the smallest acquisition time. As already explained (section 2.1), instead of choosing a set of directions and acquiring a fixed number of slices, we used a 3D sequence. Despite the reduction in acquisition time, each 3D item takes around 18 s. To keep the recording session reasonably short (actual duration was of aprox. 90 minutes), in the 3D corpus we only contemplated the sounds for which 3D can provide new important information (as for the laterals) or are reported to be somehow characteristic of Portuguese. This explains the non-inclusion of stops. For oral vowels and fricatives, only a subset of the 2D corpus was considered.

The procedures followed in this corpus were similar (excluding acquisition time) to the procedures already detailed for the 2D subset.

The corpus actual content, using the IPA phonetic alphabet [66] can be found in Table 2.

Although Alwan et al. [6] acquired sustained productions of American English rothics, EP taps and trills were not considered in this static study. We anticipated as particularly problematic to record information on [r,R] due to the several opening/closing movements involved. They have been included in a real-time MRI corpus (not presented in the paper). 3D high resolution sagittal images of the nasal and oral tracts of the speaker at rest (no phonation) were moreover acquired.

Finally, as calcified structures such as bone and teeth are not observed on MRI images, dental arches were also obtained, according to the technique described by Takemoto et al. [18], but using water as an oral MRI contrast agent. These images were however not exploited in this study and are planned to be used

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 2
2D and 3D corpus contents including target phone and reference words (in Portuguese and respective phonetic transcription using IPA phonetic alphabet) used in instructing speaker.

Phone	Word	Transcr.	2D	3D
Oral Vowels:				
[i]	pipo	[pipu]	X	X
[e]	pêca	[pekɐ]	X	X
[ɛ]	leva	[lɛvɐ]	X	
[ɨ]	devi	[dɨvi]	X	
[ɐ]	cada	[kɐdɐ]	X	X
[a]	pato	[patu]	X	X
[u]	buda	[budɐ]	X	X
[o]	tôpo	[topu]	X	X
[ɔ]	pote	[pɔtɨ]	X	X
Nasal Vowels:				
[ĩ]	pinta	[pĩtɐ]	X	X
[ẽ]	pente	[pẽtɨ]	X	X
[ɐ̃]	canto	[kɐ̃tu]	X	X
[ũ]	punto	[pũtu]	X	X
[õ]	ponte	[põtɨ]	X	X
Stops:				
[p]	[apa], [ipi], [upu]		X	
[t]	[ata], [iti], [utu]		X	
[k]	[aka], [iki], [uku]		X	
[b]	[aba], [ibi], [ubu]		X	
[d]	[ada], [idi], [udu]		X	
[g]	[aga], [igi], [ugu]		X	

Phone	Word	Transcr.	2D	3D
Nasal consonants:				
[m]	cama	[kɐmɐ]	X	X
[n]	cana	[kɐnɐ]	X	X
[ɲ]	canha	[kɐɲɐ]	X	X
Fricatives:				
[f]	fala	[falɐ]	X	
[s]	sala	[salɐ]	X	
[ʃ]	chá	[ʃa]	X	
[v]	vaca	[vakɐ]	X	
[z]	zarpa	[zarpɐ]	X	
[ʒ]	jacto	[ʒatu]	X	
[f]	[afa], [ifi], [ufu]		X	X
[s]	[asa], [isi], [usu]		X	X
[ʃ]	[aʃa], [iʃi], [uʃu]		X	X
[v]	[ava], [ivi], [uvu]		X	
[z]	[aza], [izi], [uzu]		X	
[ʒ]	[aʒa], [iʒi], [uʒu]		X	
Laterals:				
[l]	laço	[lasu]	X	X
[l]	pála	[palɐ]		X
[ɫ]	mal	[maɫ]	X	X
[ʎ]	falha	[faʎɐ]	X	
[ʎ]	palha	[paʎɐ]		X

in following studies to improve our results (see section 7.1).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

2.3 Speaker

For the 2D and 3D corpus subsets, analyzed in the present study, only one speaker was recorded (PAA). The speaker selected was an EP native speaker, male, 25 years old, 180 cm height, 70 Kg, from the north of the country, and with both vocal and singing training. The speaker had, at the time of the study, no history of speech or language disorders.

During the acquisition of all the sequences involved in the study, the speaker used headphones to respect safety recommendations related with noise levels, and also to allow for better communication. The reduced auditory feedback due to the use of headphones represents a limitation to the study, with possible negative impact on speaker's articulation.

As far as positioning is concerned, the speaker was lying in a comfortable supine position. Head and neck phased array coils were used and the speaker's head was fixed with regular foams and cushers. The speaker's head movement was later evaluated, in the 2D corpus, by analysis of the coordinates of one manually marked point supposed to be fixed in the reference coordinate system, the anterior arch of C1. Maximum movement from average (including the error of the manual marking process) was 1 pixel (corresponding to 0.78 mm) in the anterior-posterior direction and 3 pixels (2.34 mm) in the other direction. These results support our assumption that speaker's movements were negligible.

3 Image Processing

The viability of a large MRI database is determined by the existence of a reliable and fast segmentation method, with low human interaction. This is particularly relevant when using real-time MRI, where the number of images to process is very large. The study of the robustness of the segmentation method is also very important. We need to make sure that the contours generated are truthful enough to represent the vocal tract configuration of the sound being produced. The contours cannot contain errors that may lead to a misinterpretation and/or confusion of the sound with another one. This can be evaluated with a metric called the Pratt Index [67].

All image analysis operations were performed in Matlab, version 7.0.1. The code used was specially implemented by one of the authors for use in this work. Exception is made for the `live wire` routine, developed by Chodorowski et al. [68]. We were able to obtain 2D contours, articulatory measures, area functions, quantification of the velum port opening, and 2D/3D visualizations

ACCEPTED MANUSCRIPT

of the vocal tract. To achieve these goals, the image analysis process included mainly: (1) 2D segmentation of the vocal tract, (2) 3D segmentation of the vocal tract and area extraction of the sections, and (3) computation of the velum port opening quotient (VPOQ).

3.1 2D Corpus

The 2D segmentations were made with the *region growing* method [69]. We started by manually placing a seed inside the vocal tract which expanded until it reaches the vocal tract wall. This expansion is based on grey level comparison between the mean grey level value of all the pixels already marked as inside the vocal tract and the neighbour pixels of the contour of the region already delimited. The stop criterion is based on a maximum difference threshold between the pixel being tested and the mean value of all the pixels assumed to belong to the region of interest.

To assess reproducibility of the process, 100 contours were generated (each set takes about 35 minutes with the current implementation) with a randomly placed seed inside the vocal tract, for each image. Each contour was compared with the mean contour (chosen as reference contour). Comparisons between contours were made with the Pratt Index (abbreviated as PI) [67], a distance between two contours defined by: $PI = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{1+\alpha d_i^2}$, where N is the number of corresponding points between contours, d_i is the distance between two corresponding points, and α is related to the contour size. Based on one of the authors' previous work on other types of images [67], $\alpha = 1/9$. Corresponding points between contours are obtained as follows: first contour with the smaller number of points is chosen; for each point of this contour, the closest point in the other contour is the correspondent point. This index has its range in the interval [0, 1], where 1 means that the two contours are equal. The PI was also used to compare images of different sounds. In this case, we retained 101 PIs for each pair: the PI calculated between the two mean contours (resulting from the process described above) and the 100 PIs resulting from comparison of the contours corresponding to different seeds. As no order effect was anticipated, the 100 contours for each sound were compared with the contours of the other image by their order of calculation.

Fig. 1, presents, separately, the results obtained for oral vowels, nasal vowels and consonants, showing that the *region growing* segmentation method is robust to changes in the seed (low intra-variability). The corresponding PIs are close to 1, having as a minimum the value 0.84.

Also interesting, for validating the process, is the comparison between the PIs calculated for the contours obtained for one sound (intra-variability) and

ACCEPTED MANUSCRIPT

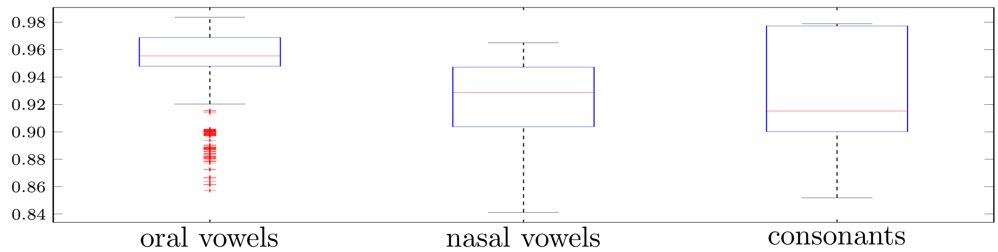


Figure 1. Boxplots of the Pratt Index differences obtained by using different starting points (seeds). Results for oral vowels, nasal vowels and nasal consonants are presented.

the PIs obtained for different sounds (inter-variability). Fig. 2 presents these results.

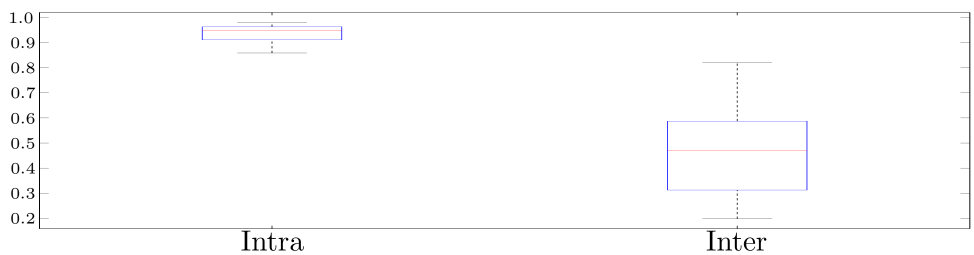


Figure 2. Boxplots comparing Pratt Index of all contours obtained with different starting points for a fixed image (Intra) and contours of different EP sounds (Inter). In the calculation, part of 2D corpus was used: all oral vowels, all 5 nasal vowels and the consonants [m], [s] and [l].

The 95% confidence intervals [70,71], calculated using SPSS, are: $CI_p[0.92 \leq Intra \leq 0.96] = 95\%$ for the intra-variability, and $CI_p[0.44 \leq Inter \leq 0.49] = 95\%$ for the inter-variability, resulting in a statistically significant difference between the variability due to the segmentation starting points and the differences due to different sounds.

All 2D sagittal images were also manually marked with the following relevant points (Fig. 3): highest position of tongue dorsum (TD); tongue tip (TT); tongue root position at the C3-C4 vertebral level (TR); jaw height, using the root of lower incisors (JH); lower lip highest and most anterior position (LL); and upper lip lowest and most anterior position (UL). TR is the intersection with tongue contour of an horizontal line passing through C3-C4 level. Note that all TR measures have therefore the same vertical coordinate value and that the discrepancy observed on Fig. 7b is around 1 mm and can be ascribed to the general process accuracy. We used as origin the lower left image point, and assumed that the speaker movement is not relevant. A different reference point could easily be chosen.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

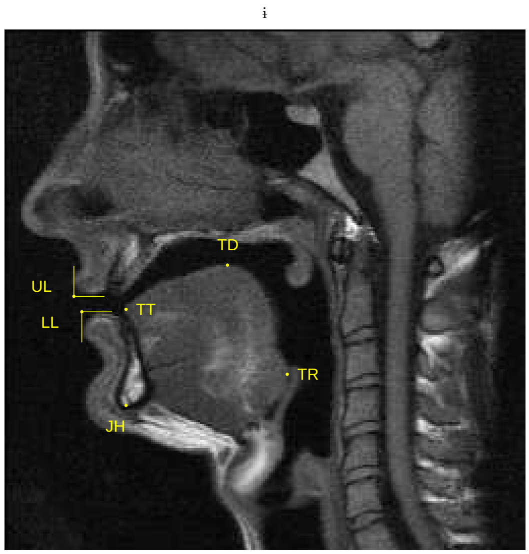


Figure 3. Midsagittal profile obtained during the production of a sustained [ɨ] by PAA, as in the word (devi) [dɨvi], showing measured articulatory points. Articulatory points used for this work are: highest tongue dorsum point (TD), tongue tip (TT), tongue root position at C3-C4 level (TR), jaw height (JH) and lower (LL) and upper lip (UL) spatial coordinates.

3.2 3D Corpus

For the volumes, we first segmented the vocal tract in the midsagittal slice using the semiautomatic technique *live wire* [68]. Next a (fixed) gridline was applied and its intersections with the contour obtained. Middle points between the intersection in the 2 contour parts make our first approximation to the centerline. The centerline is then upsampled and smoothed. Then the volume was resliced according to a phoneme-adapted grid with planes oriented normally to the centerline. Each slice was also segmented using the live wire technique. We opted to use a number of slices similar to the used in other studies, 45 slices, covering all the oral tract. Although having a non-isotropic voxel, which is homogenized by a linear interpolation, we believe that with this method we will obtain more realistic data.

The live wire segmentation approach is based on optimal search strategies over graphs built upon regional pixel maps defined on the neighbourhood of seed points determined by the user. This is a fully semiautomatic approach taking advantage of the unsurpassed human capacities for object recognition and delineation. Typically, the user starts segmentation by choosing an initial point (seed) on the boundary of the object of interest. Then, the algorithm computes the minimal cost path between the seed and the current position of a pointing device (mouse pointer). The criterion for minimal cost is often the integral of pixel intensities along a path. This minimal cost path is rendered

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

continuously (the live wire paradigm) as a partial contour and if the user considers this partial contour as acceptable then he can proceed and define the next seed point. After a minimum set of seed points the boundary of the target object, not necessarily closed, should be completely delineated. Relying on the user pattern recognition capabilities, the live wire approach offers a sequence of locally optimal contours and it is often the segmentation technique of choice to deal with difficult images with diffuse targets and cluttered backgrounds. This segmentation technique was adopted due to its better performance in the lower image quality of the 3D resliced images, when compared with the *region growing* technique used for 2D Corpus

As can be observed in Fig. 4, each resliced plane will have an orientation perpendicular to the centerline of the vocal tract. The bottom part of the vocal tract is usually easy to segment in these resliced planes, but some difficulties were found in the segmentation of the oral cavity.

For validation purposes, a sample of the 3D segmentations was visually evaluated by two experts

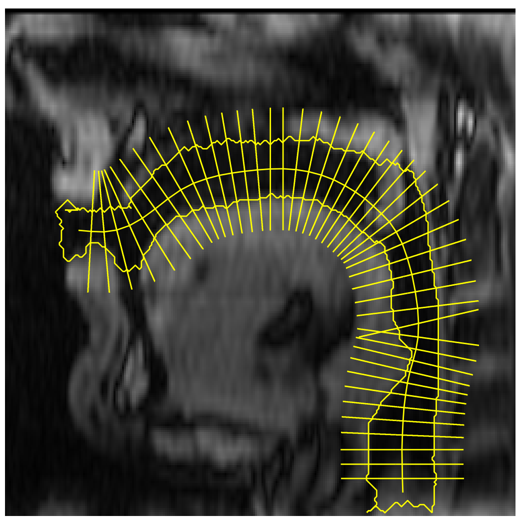


Figure 4. Example of a resliced midsagittal cut, for [a], obtained from the volumetric information (between a few centimeters above hard palate to C5 vertebral level). Superimposed, the generated adaptative grid is shown. With this procedure all obtained slices are orthogonal to the vocal tract centerline.

Difficulties in observing larynx area, due to 3D aliasing, motivated the use of a reference point for our area functions at the basis of C5 vertebral body. Thus, in the obtained area functions, x-axis represents the distance from this reference level towards the lips, representing 0 the basis of C5 and not the larynx position. As the basis of C5 was marked separately from the process of area function determination, it is possible that area function started after this reference point. We also did not put much effort into improving segmentation of this lower part of the pharynx, not forcing the centerline to go as close

ACCEPTED MANUSCRIPT

as possible to the larynx position. We prefered to concentrate on the other parts of the area functions. However, this imprecision around glottis should be improved in the future, leading to more accurate area function lengths.

The **VPOQ** was computed in a similar way to [56]. In this method, we identified the first slice (from the glottis to the lips) where both the oral and nasal cavities can be seen. We then chose that slice and the next four and measured the area of the oral and nasal passages. Mean VPOQ was calculated as the mean of the quotients between the nasal and oral areas, for the 5 slices. In Fig. 5 the first oblique slice is shown (counting from the glottis to the lips) where both the oral and the nasal cavities are visible.

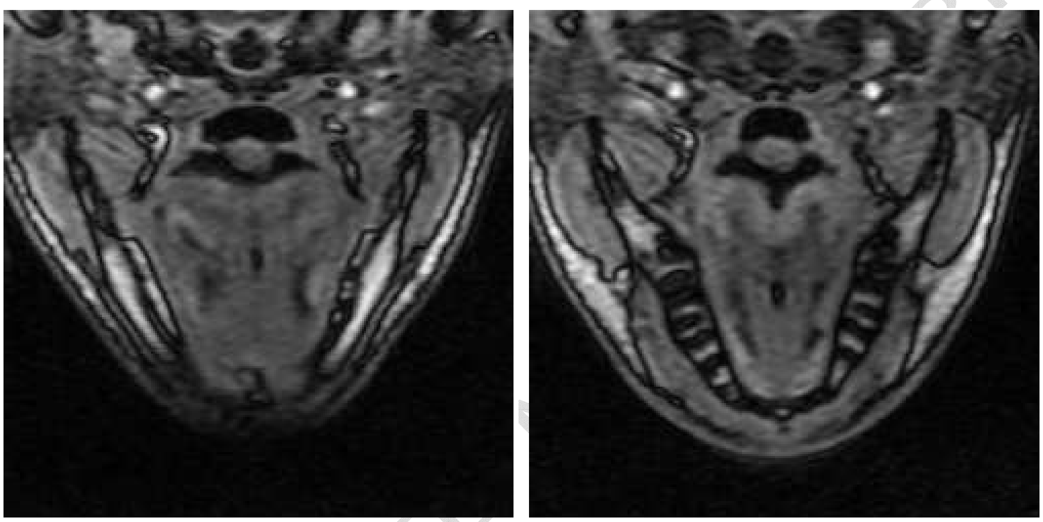


Figure 5. Examples of coronal oblique views obtained from nasal consonants 3D data: [m] at left and [ɲ] at right. The cut passes through the velum (orthogonal to the vocal tract centerline). Two passages can be observed: one (at the top) refers to nasal cavity and the other to oral cavity (bottom).

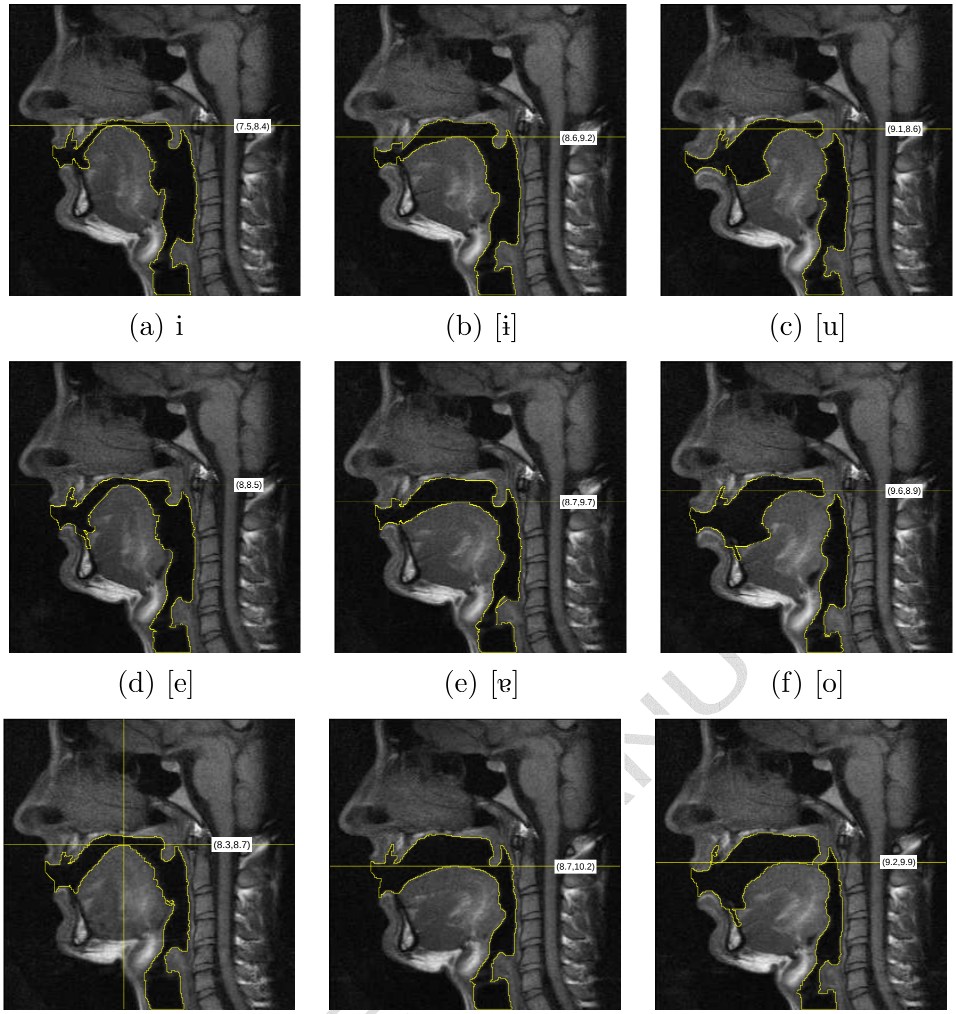
4 Results I: Vowels

We start this study with the analysis of the oral vowels. After we present our results for nasal vowels. At the end of the section a comparative study of nasal and oral vowels is also presented.

4.1 Oral vowels

We present the MRI images with superimposed contours for the 9 oral vowels in Fig. 6. Vowels are arranged according to their phonetic description, high

ACCEPTED MANUSCRIPT



(a) i (b) [ɨ] (c) [u]

(d) [e] (e) [ɐ] (f) [o]

(g) [ɛ] (h) [a] (i) [ɔ]

Figure 6. Midsagittal images with superimposed contours for the EP oral vowels: from the top, [i], [ɨ], [u], [e], [ɐ], [o], [ɛ], [a] and [ɔ].

vowels at the top and posterior vowels to the right (in agreement with orientation of our images, with lips to the left). The corresponding articulatory measures (TD, TR, TT, JH, UL and LL) are presented in Fig. 7. The area functions are presented, separately, in Fig. 8. The following descriptions were based on all the information available, particularly in the parameters presented in Fig. 7.

4.1.1 Anterior oral vowels

Regarding the tongue highest point (TD), [ɛ] is produced with the lowest position of TD; [i] with the most raised and anterior position; [e] in an intermediate position in both dimensions, being closer to [ɛ] in the anterior-posterior axis.

Looking at the [i] and [e] area functions, Fig. 8, (corpus doesn't include 3D for [ɛ]) the point of smallest area is more anterior for [i], confirming TD parameter information. In the area functions it is possible to see that for [i] the constricted area is a few centimeters long while in [e] the obstruction zone it is much more restricted.

It has also been observed, that the most posterior tongue position (TR) is more anterior in [i] than in [ɛ], contributing to the increase of the pharyngeal cavity and the reduction of the oral cavity. The wide pharyngeal region for [i] is indeed clear on area functions.

The JH is lower in [ɛ] and higher and more anterior in [i].

The TT vertical position increases from [ɛ] to [i], being [e] closer to the [i]. The distance between [e] and [ɛ] is almost twice the distance between [e] and [i]. In the horizontal direction differences are smaller: [i] and [ɛ] present very similar TT horizontal positions; [e] has a slightly posterior position.

Regarding lip configuration, the results are different for the upper and lower lip. The 3 anterior vowels present quasi-identical UL parameter values. For lower lip (LL): [i] presents a higher position; protrusion (x-axis position) is not very different for the 3 vowels; differences are mainly in the vertical position, being [i]-[e] and [e]-[ɛ] distances similar.

For each one of the three configurations, the velum is raised, not having a significative position alteration among the three vowels. In the region of the glottis there is no evidence, in the sagittal plan and for this speaker, of alterations between the three vowels.

In terms of the similarity of contours, with the analysis of PI, [e] is closer to [i] (PI=0.76) than to [ɛ] (PI=0.72). Despite the very similar values of PI in both cases, non-parametric statistical tests (Mann-Whitney) confirm the difference as significative ($p < 0.001$).

4.1.2 Central oral vowels

The vowel [ɨ] (high vowel) is produced with the tongue dorsum (TD) in the highest position inside of the series; followed by [ɐ] and [a] (low vowel). All 3 have similar x-coordinates for TD. Comparing with the anterior vowels, TD is always lower for central vowels. The highest value for central vowels (10.9) is clearly lower than the lowest position for anterior vowels (11.3). For this series of vowels, TD is not directly related with maximum constriction position, area function provides further insight. Our data show [a] as having its smallest area in the pharyngeal region.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

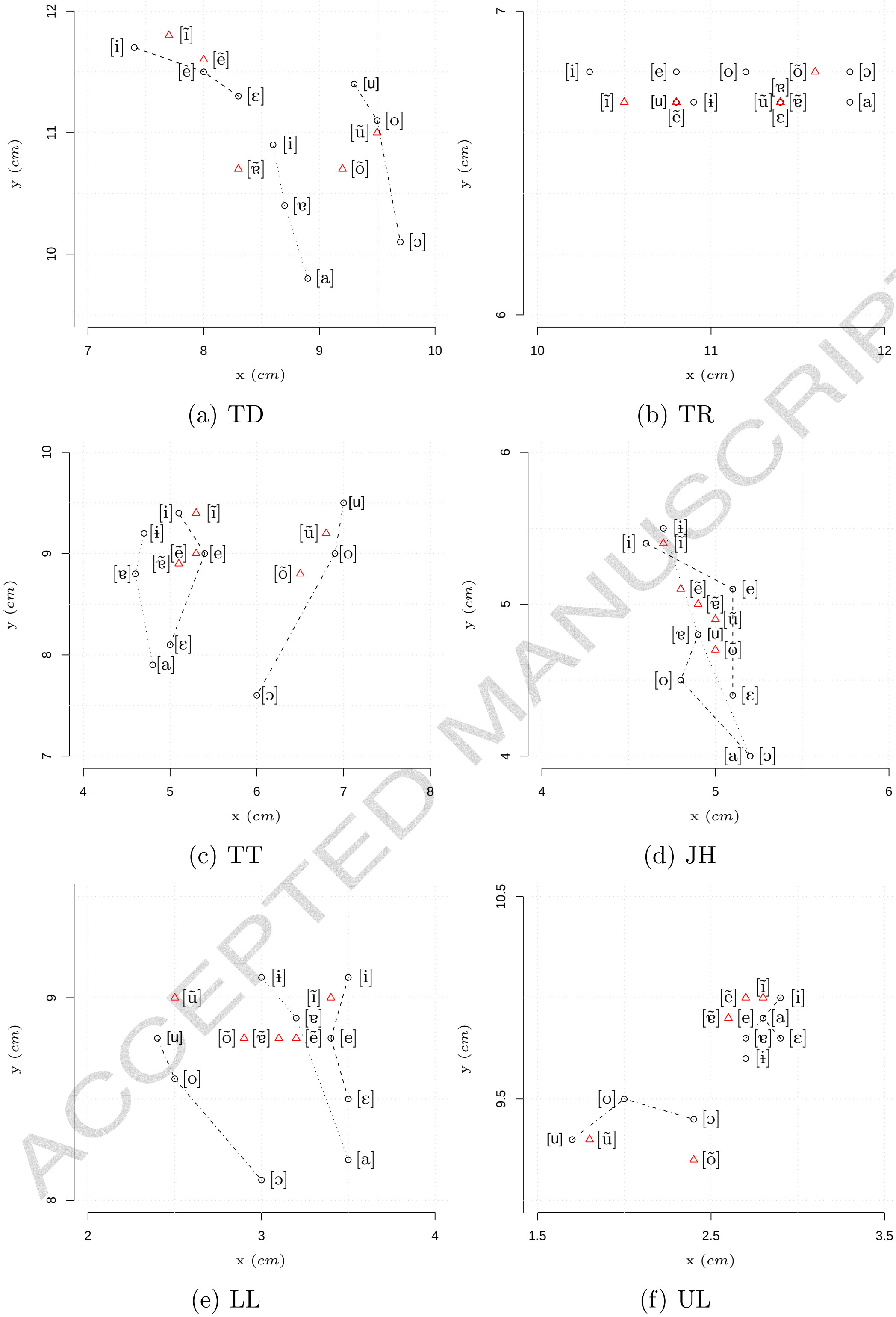


Figure 7. Six articulatory measures for EP vowels. From the top left: Tongue dorsum highest position (TD), tongue root at C3-C4 level (TR), tongue tip (TT), jaw height (JH), and lower (LL) and upper (UL) lip.

ACCEPTED MANUSCRIPT

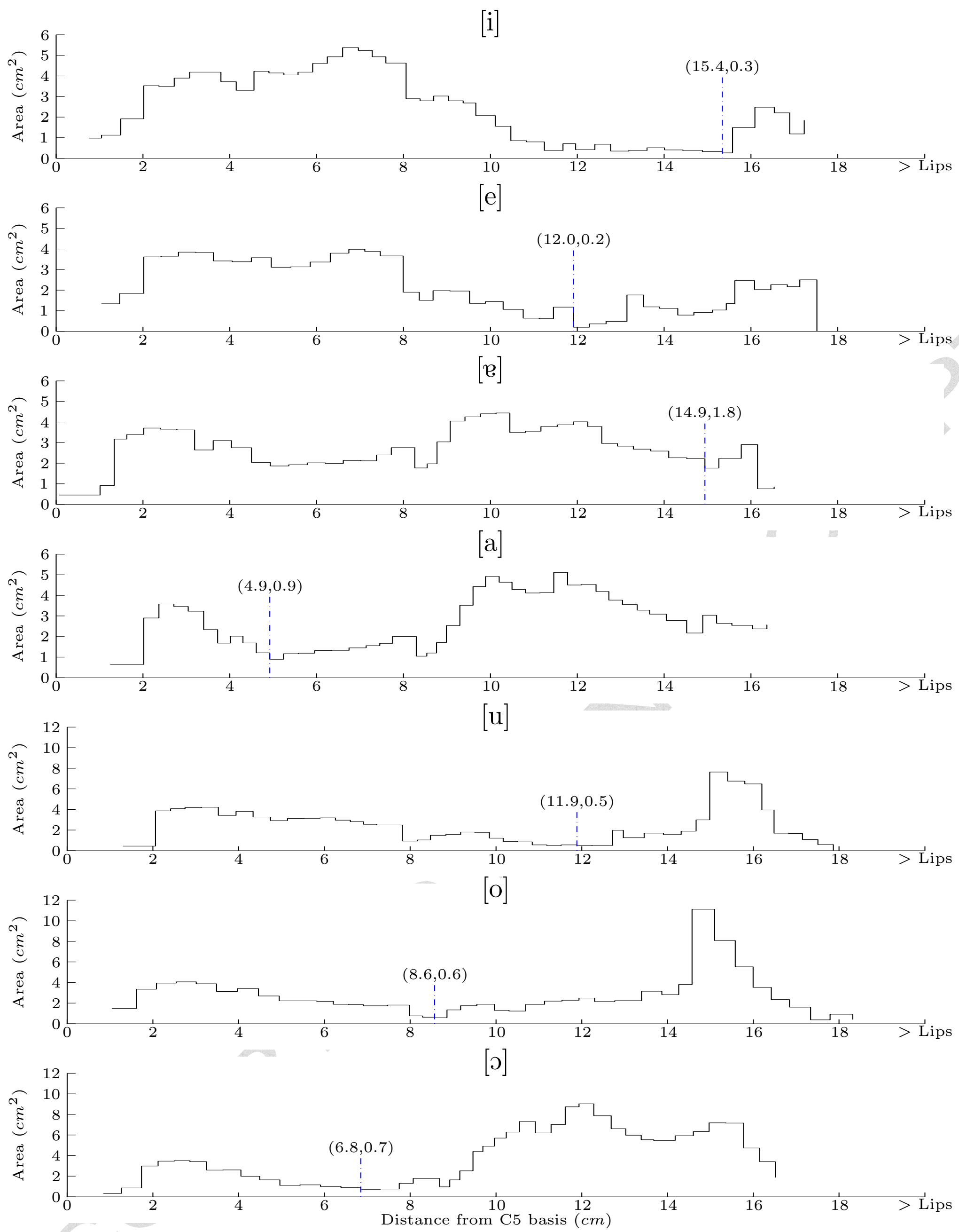


Figure 8. Area functions for seven of the EP oral vowels. They were grouped in anterior, central and posterior, with higher vowels at the top. From the top, [i], [e], [ɐ], [a], [u], [o] and [ɔ]. In the area functions, information regarding the constriction point (distance from reference, at basis of C5 vertebra, and area) is included. Note the difference in y-axis scale for the three last area functions, with a maximum twice of what was used in the others.

The tongue root (TR) is more anterior during the production of [ɨ] than of [ɐ] or [a]. [a] is also more posterior than all 3 anterior vowels. In terms of area function, major differences between [a] and [ɐ] are in the pharyngeal region.

The jaw position (JH) is lower and posterior for [a] and higher and anterior for

ACCEPTED MANUSCRIPT

[ɨ]. There is an overlap of the opening values with anterior vowels. Nevertheless, [a] is produced with the lowest position in the combined anterior-central set of vowels.

The tongue tip (TT) position follows the same pattern observed for TD, with a correlation between the points.

The lower and upper lips positions can be considered as nearly similar for [ɐ] and [ɨ]. In [a], the lower lip is lower, around 7 mm, and, also, more posterior (5 mm). This may be related to mandibular position.

From contours superimposition, not shown in this paper, the velum presents a more anterior position in the vowels [ɐ] and [ɨ] than in [a].

Non-parametric statistical tests (Mann-Whitney) showed: as non-significantly different the PIs obtained for the comparisons of [ɐ] with [ɨ] and [ɐ] with [a]; as significantly ($p < 0.001$) higher the similarity of these two comparisons than similarity between [ɨ] and [a].

4.1.3 Posterior oral vowels

It can be observed in Figures 6 and 7 that vowel [u] is produced with the highest TD position amongst the three posterior vowels, followed by [o] and [ɔ], with the lowest and more posterior position. Compared to anterior and central vowels, posterior vowels are produced with lower TD than the anterior series. Only [a] is produced with lower TD than the lowest posterior ([ɔ]), and only with 3 mm difference. When compared with anterior vowels we observe that posterior vowels have, generally, lower TD position, except for [ɛ], which is slightly lower than [u]. Comparing posterior and central vowels, it can be observed that TD for [u] and [o] is higher than the value for the three central vowels. In the area functions, the point of maximum constriction follows the same tendency of TD parameter to lower from [u] to [ɔ], moving downward in the pharyngeal region. Tongue root position on sagittal images also confirms a more posterior position for [ɔ] than for [u] and [o]. The difference between [u] and [ɔ] is about 1 cm. The tongue back position is closer to the velum in [u] and [o], while in [ɔ] is directed towards the pharyngeal wall. This dorsovelar orientation for [o] was an unexpected finding since this oral vowel is generally described as being produced with tongue back oriented towards the pharynx (e.g. [72, p. 53]). From midsagittal profiles, corroborated from area functions values, an increase of oral cavity dimension from [u] to [ɔ] is evident, associated with a decrease of the pharyngeal cavity dimensions.

Comparing TR positions for anterior and posterior vowels (Fig. 7b) we can observe a trend for anterior vowels to have more anterior TR positions, but with an overlap of the two classes (e.g. [ɛ] is more posterior than [o]).

ACCEPTED MANUSCRIPT

The jaw (JH) is lower in the production of [ɔ] than in the production of [o] and [u], these two vowels being produced with JH respectively 5 mm and 8 mm above. For tongue tip (TT) we notice a similarity between [u] and [o], both with TT more posterior and higher than [ɔ]. When comparing with the 2 previous series, in posterior vowels the range of values for TT is larger, both in the horizontal and vertical dimensions. While for central and anterior vowels TT has a maximum range of 0.4 cm in the horizontal and 1.3 cm for vertical, the ranges are 1.0 cm and 1.8 cm for posterior vowels. Also relevant to this series is the variation of lip position, particularly protrusion. Protrusion is important for [u] and [o]. For [ɔ], lower lip protrusion is smaller and similar to the highest value obtained in previous series (for [ɨ]). When compared with anterior and central vowels, the difference is marked, as expected, since in EP only posterior vowels are rounded.

From the superimposition of contours (not included in the paper), it can be observed that the velum is in a lower position in the production of [ɔ] than in the other two posterior vowels.

Area functions for [u] and [o] present a similar pattern, contrary to [ɔ]. Pattern differences are more pronounced at oral cavity level. Analyses of the PI, confirm this tendency, as PI between [u] and [o] mean contours is 0.77, being 0.73 between [o] and [ɔ], and 0.65 between [u] and [ɔ]. Statistical tests (Mann-Whitney) confirm as significantly higher the values of the PI for the pair [u] and [o] when compared with both other two pairs ($p < 0.001$).

4.2 Nasal vowels

Fig. 9 show the images with superimposed contours and area functions for EP nasal vowels, complementing the information presented in Figs. 7 and 10. Based on this 3 Figures, we can observe that:

- Vowels [ĩ] and [ẽ] are produced with the tongue (TD) in an anterior and raised position.
- Vowel [ɐ̃] has a low TD position, occupying with [õ] the lowest TD positions measured for the 5 nasal vowels.
- Vowels [õ] and [ũ] are more posterior in terms of TD.
- The jaw position, in contrast with what happens in the production of the oral vowels, presents a more restricted range of variation. For the five nasal vowels higher and lower JH measures differ of 0.7 cm while for oral vowels difference is more than the double, 1.5 cm.
- The velum is open for all nasal vowels, but its height is variable with the vowel. We will study these differences, below, using 3D information.
- Labial protrusion is marked in the production of [ũ] and similar to the

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

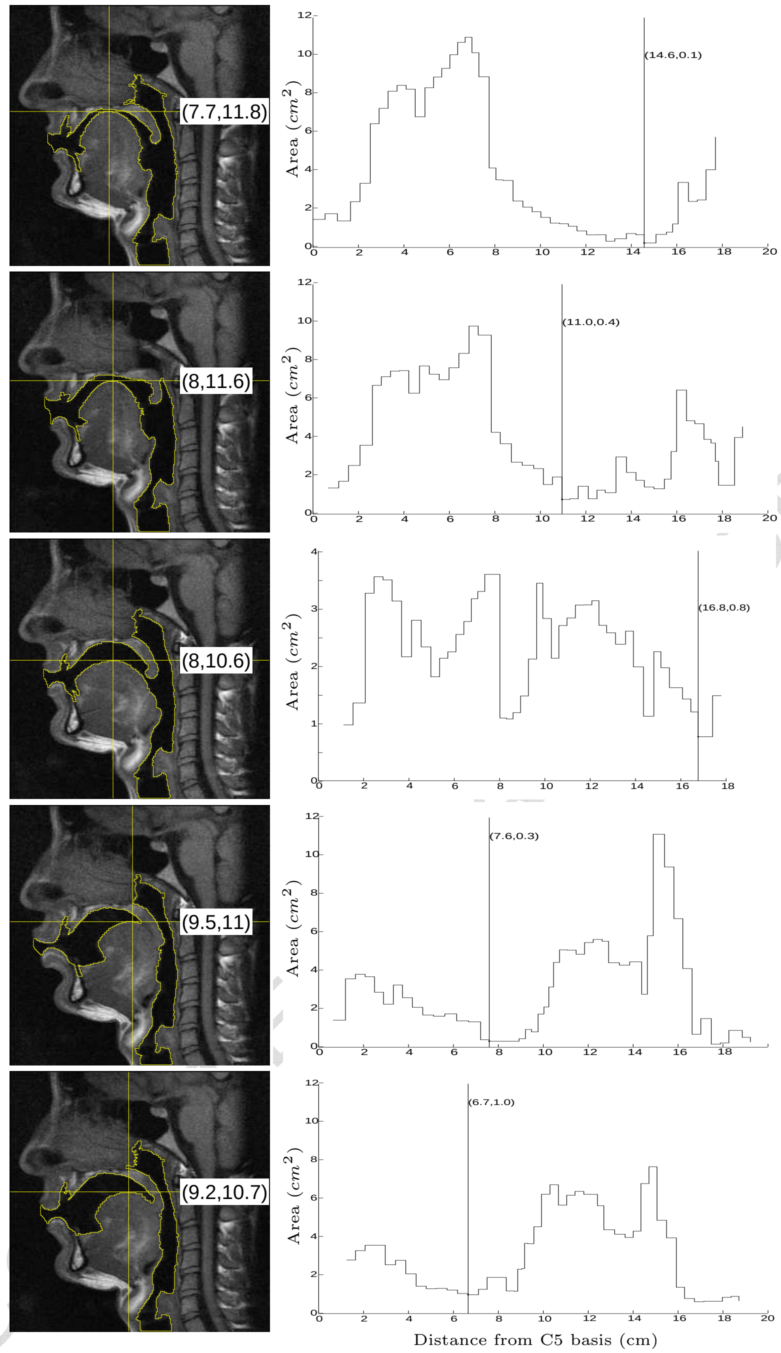


Figure 9. Results for the 5 EP nasal vowels: from the top, [ĩ], [ẽ], [ɐ̃], [ũ] and [õ]. In each row, are presented, from left, the midsagittal image with superimposed contour and area function. In the area functions, information regarding constriction point (distance from reference point and area) is included.

581 protrusion observed in the corresponding oral vowel ([u]).

ACCEPTED MANUSCRIPT

4.2.1 Nasal vs. Oral vowels

In this subsection comparisons between oral and nasal vowels are presented. They are based on the articulatory measures of Fig. 7, the superimposition of midsagittal contours for EP nasal vowels with their possible oral counterparts (Fig. 10) and area functions obtained from 3D acquisitions (Fig. 11). For mid and low nasal vowels two oral configurations are considered.

With MRI 3D information we can, for the first time for EP, compare the area functions of oral and nasal vowels. Differences between two area functions were obtained as follows: both area functions were resampled at the same positions along the x-axis, resulting in two vectors with the same length; the difference is the result of subtracting the two vectors.

The vowels [ĩ] and [i] present similar configurations, the nasal vowel being produced with a higher and posterior position of the tongue body and root when compared with the oral counterpart (Fig. 10(a)). The TD position is close for the two vowels, being (7.4 cm, 11.7 cm) for the oral and (7.7 cm, 11.8 cm) for the nasal (Fig. 7a).

The nasal [ũ] is produced with a slightly posterior and lower TD than the oral counterpart [u] (Fig. 7a). Looking at Fig. 10(b), comparison of [e], [ɛ] and [ẽ], we can observe that the contours of the vowels [e] and [ẽ] are closer (PI= 0.86) than the contours of [ɛ] and [ẽ] (PI=0.69). Specifically with respect to TD position, the nasal vowel [ẽ] is produced with the highest TD (Fig. 7a), this difference being however more accentuated for [ɛ] than for [e]. The oral [e] and the nasal [ẽ] present a similar pattern at pharynx level, which is not valid to [ɛ], more constricted than [ẽ]. Differences at tongue tip level (TT) are small between [e] and [ẽ] and more pronounced between [ɛ] and [ẽ]. The velum although opened during the production of the nasal, seems to be in a higher position than in the other nasal vowels. This tendency is observable in contours superimposition not included in the paper. From 3D information (only relative to [e] and [ẽ]), we confirmed that the nasal and the corresponding oral vowel [e], have a very similar pattern on area function.

Analyzing Fig. 10(c), we can detect some differences. The nasal vowel [ɐ̃] is produced with a TD in a higher position than for [ɐ] and [a]. In the anterior-posterior axis, [ɐ̃] has a TD more anterior than all 3 EP central oral vowels, in a position similar to anterior oral vowel [ɛ]. The tongue root (TR) is similar for [ɐ̃] and [ɐ] and more posterior for [a].

Observing Fig. 10(d), we detected that, with respect to tongue height, the nasal vowel [õ] is produced between [o] and [ɔ]. In the tip/blade region, and looking at the TT parameter, the configuration of [õ] is closer to [o] than to [ɔ]. Regarding TR, [õ] is between [o] and [ɔ].

ACCEPTED MANUSCRIPT

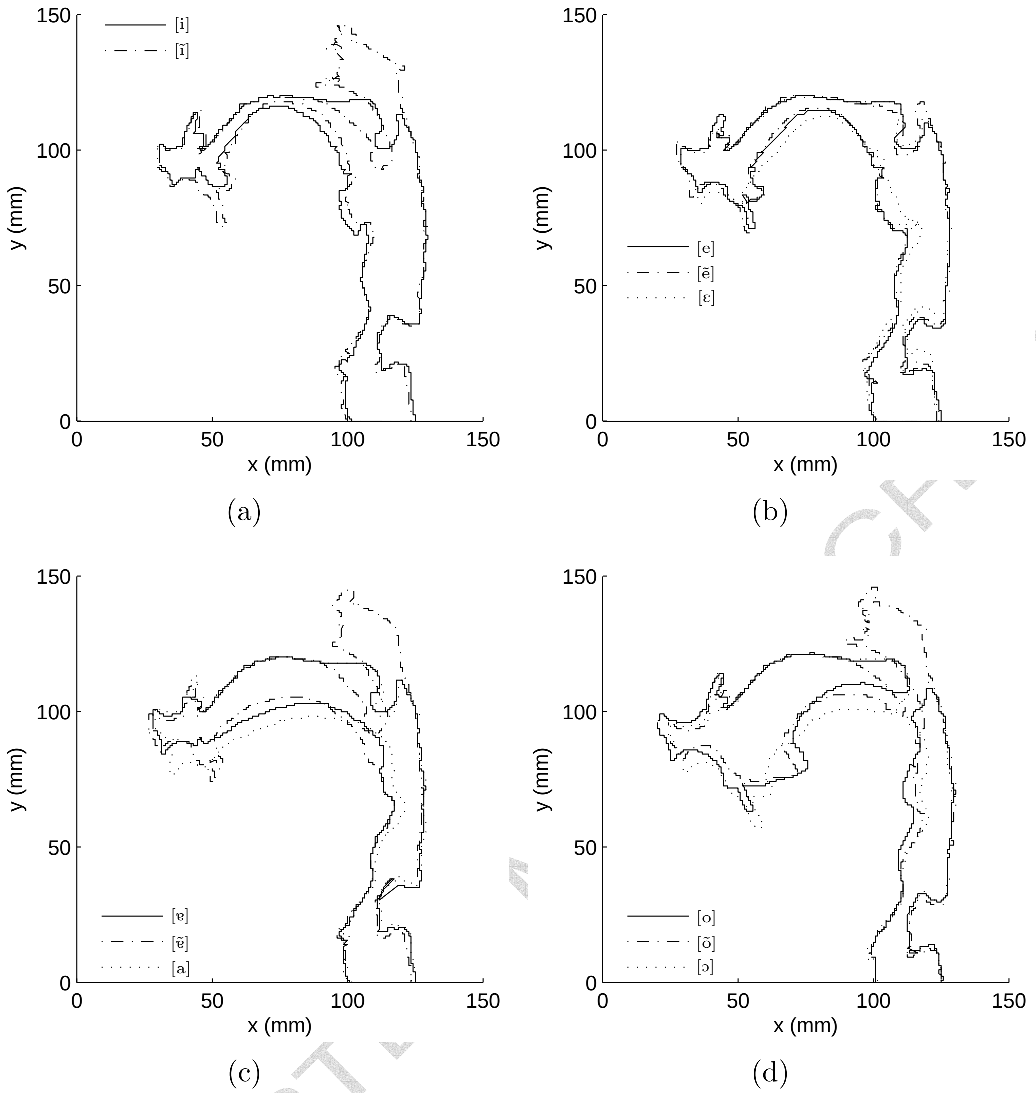
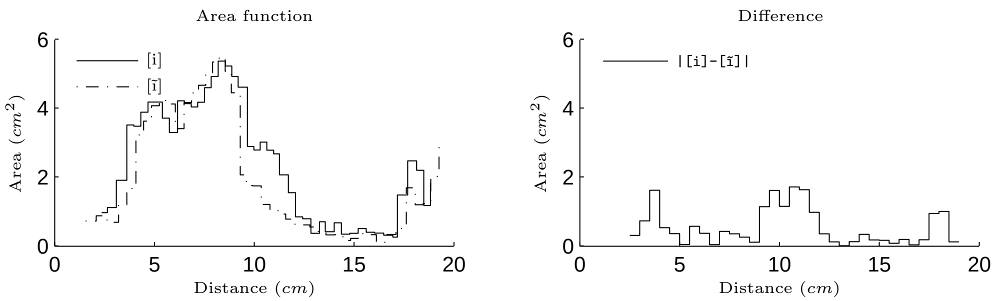


Figure 10. Midsagittal vocal tract profiles comparisons for nasal vowels and their possible oral counterparts: (a) superimposition of [i] (solid line) and [ĩ] (dash-dotted); (b) superimposition of [e] (solid line), [ẽ] (dash-dotted) and [ɛ] (dotted); (c) superimposition of [a] (solid line), [ɐ̃] (dash-dotted) and [ɐ] (dotted); (d) superimposition of [o] (solid line), [õ] (dash-dotted) and [ɔ] (dotted).

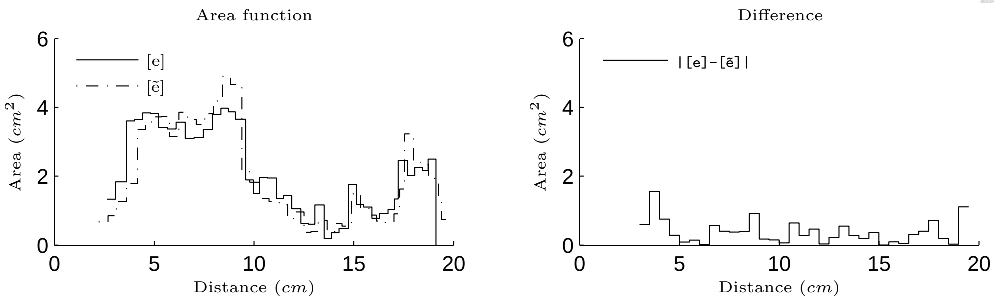
In these midsagittal images it is apparent that velum and uvula touch the tongue back during the production of back vowels [õ] and [ũ]. For the other nasal vowels this is not observed.

Midsagittal distances in the pharyngeal cavity are different in nasal vowels and their oral counterparts. As an example, [ɐ̃] has a wider upper pharynx region relative to [ɐ]. During the production of EP oral and nasal vowels, there are not noticeable differences with respect to posterior wall of the pharynx.

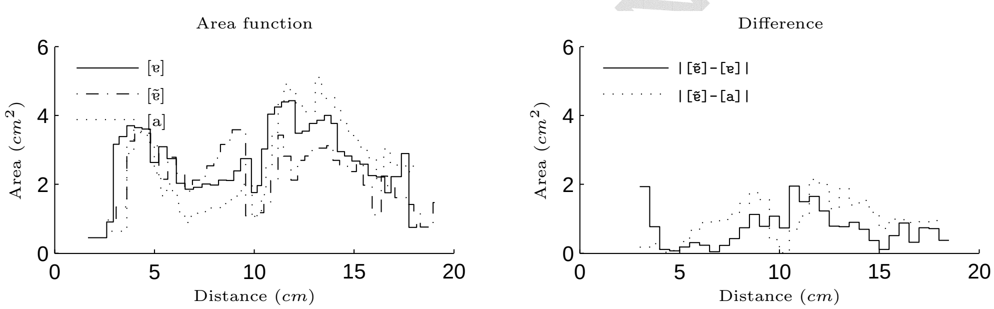
ACCEPTED MANUSCRIPT



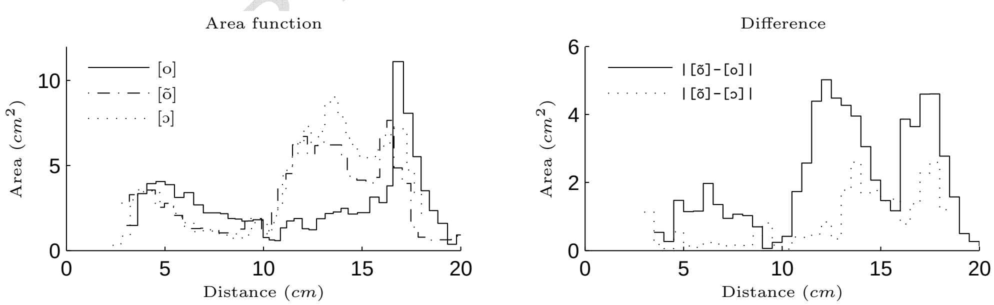
(a) Comparison between [ĩ] and [i].



(b) Comparison between [ẽ] and [e].



(c) Comparison between [ɐ̃], [ɐ] and [a].



(d) Comparison between [õ], [o] and [ɔ].

Figure 11. Area functions comparison between EP nasal and oral vowels. On the left, a plot of area functions; on the right the absolute differences between nasal vowel and oral counterparts.

4.2.2 VPOQ

A particularly interesting parameter to study for the nasals is the VPOQ. The results obtained for EP are presented in Fig. 12. We can observe that:

ACCEPTED MANUSCRIPT

- for this speaker, the average VPOQ is always higher in the nasal vowels than in the corresponding oral ones;
- [ɐ̃] presents the highest VPOQ, followed by [ũ] and [õ];
- the largest oral/nasal VPOQ difference was observed in the pair [ɐ]/[ɐ̃];
- the smallest oral/nasal difference is between [u] and [ũ].

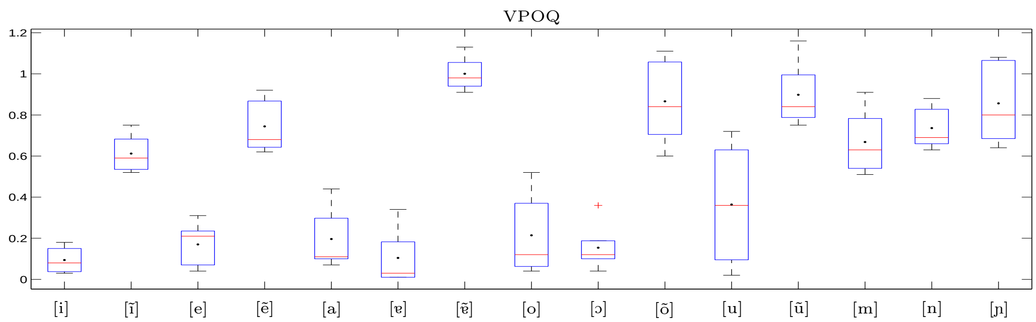


Figure 12. Boxplots of VPOQ for oral vowels, nasal vowels, and consonants. Dots represent the VPOQ average value.

5 Results II: Consonants

In this section, relative to consonantal sounds, we start with the description of the nasal consonants, to maintain continuity with the anterior section on nasal vowels. Next, stop consonants are briefly described as they are not generally considered as significantly different from other languages. They follow nasal consonants to allow a comparison between these two related classes. Then, we present results concerning fricatives, ending with a class with some EP particularities, the laterals. As the consonants depend on vocalic context, we are limited in the description of articulatory differences. Despite the use of similar vocalic context in the words used to instruct the speaker for the non VCV parts of the corpus (in general an [a] follows the consonant), we avoided descriptions that could be more related to the production of the vowel than to the consonant we are studying.

5.1 Nasals

In Fig. 13 midsagittal MRI images, contours and area functions for the EP nasal consonants are presented. In Fig. 14 a comparison between EP nasal and stop consonants contours is presented.

In these images, the different places of articulation and the open position of the velum are clearly visible. The nasal [m] is produced with lip closure, [n] is produced with tongue tip occlusion at the superior incisors, and [ɲ] is clearly produced with tongue touching the hard palate.

ACCEPTED MANUSCRIPT

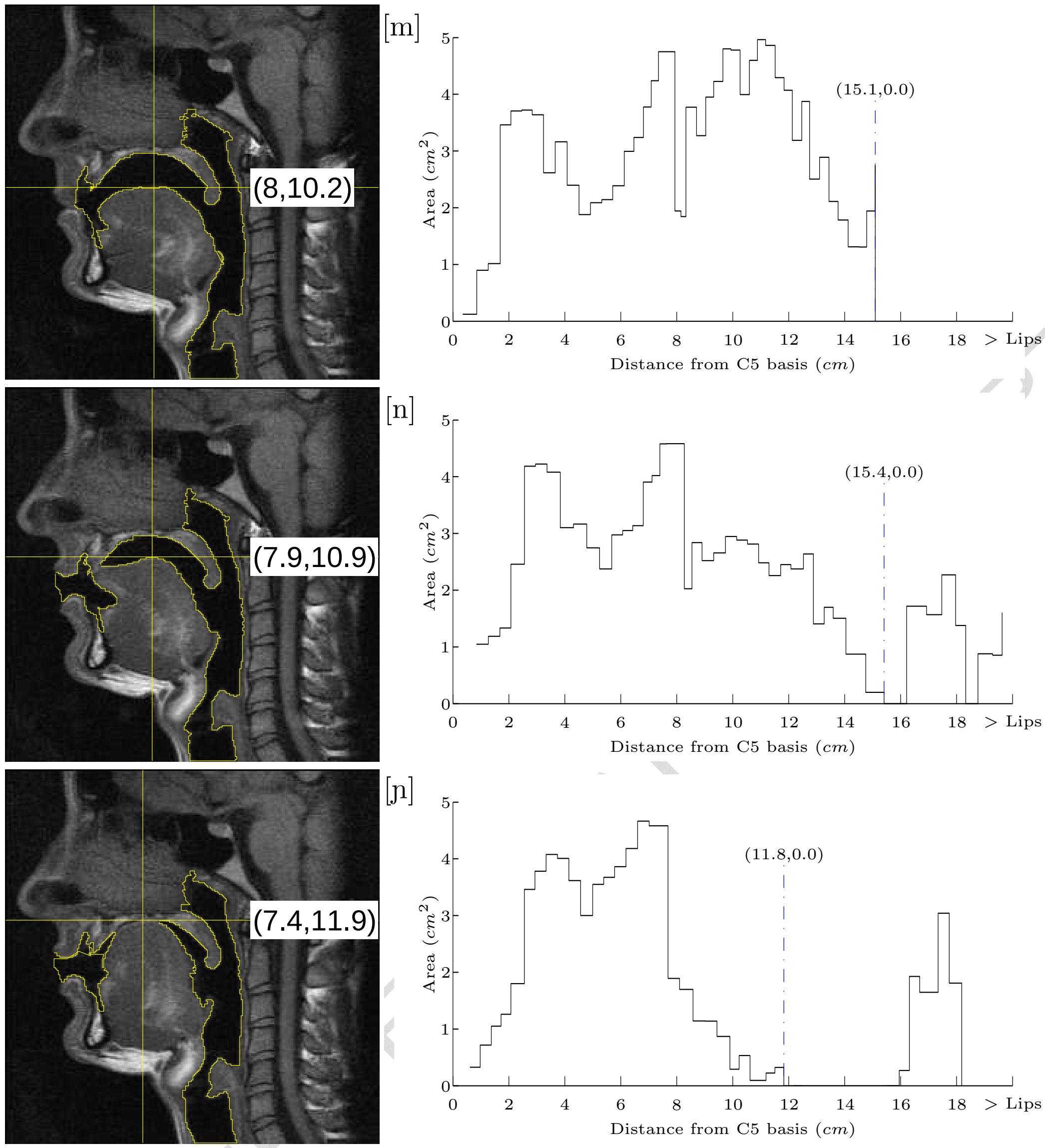


Figure 13. Results for the EP nasal consonants. From the top, bilabial [m], dental [n] and palatal [ɲ]. All the 3 sounds were sustained having a reference word with the same symmetric vocalic context, the oral vowel [ɐ]. In each row the following are presented: the image with superimposed contour (at left) and area function. In the area functions, information regarding occlusion point (distance from reference point and area) is included.

The tongue dorsum's highest point (TD) is more anterior for [ɲ] being similar for [m] and [n]; higher, as expected, for [ɲ], followed by [n] and finally [m]. [ɲ] is only 1 mm higher than [ĩ] and 2 mm higher than [i], the highest vowel TD.

The tongue tip (TT), involved in the articulation of [n] and affected in [ɲ] due to the overall raised tongue configuration, obviously presents very different positions.

ACCEPTED MANUSCRIPT

Looking at the contour comparisons for nasal consonants and stops with the same place of articulation, in Fig. 14, the main differences occur in the (upper) pharyngeal region with a more forward position of the tongue root for nasal consonants, associated with a lower position of the velum. EP stops have a narrower pharynx when compared with nasal consonants. This difference is more noticeable in the dentals ([n] vs [t]) than in the bilabials ([m] vs [p]). For the same place of articulation, nasal consonants present a more constricted larynx than stop consonants.

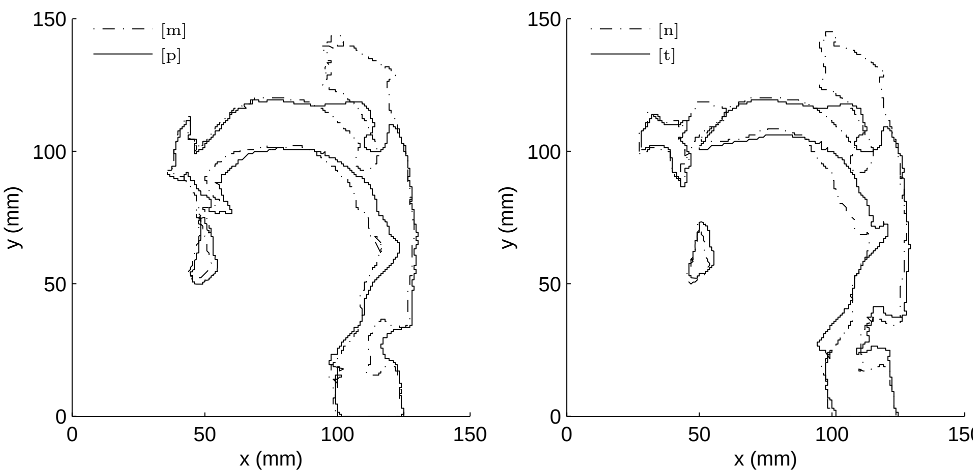


Figure 14. Midsagittal contour superimposition for nasal consonants and stops with the same place of articulation. At the left, bilabials [p] and [m]; at the right the dentals [t] and [n]. The 2 nasal consonants were sustained having an example word with the same symmetric vocalic context, oral vowel [ɐ]. The stops are the ones produced in the [aCa] context.

VPOQ for nasal consonants was already included in Fig. 12. Nasal consonants present, on average (mean=0.75), intermediate values between the nasal vowels (mean=0.82) and oral vowels (mean=0.19).

5.2 Stops

In Fig. 15, left column, we can verify that in the production of [p] there is lip closure, as expected for a bilabial stop. In the production of [t] (although teeth contour is not visible) we see an approach of the tongue tip to the dental region. In the production of [k], the articulation point does not seem clearly velar, the constriction being in the transition between the palate and the velum.

Also in Fig. 15, right column, we can observe that voiced stops present configurations that are close to the unvoiced, sharing the same articulation point. This was confirmed by contour superimposition and calculation of mean dif-

ACCEPTED MANUSCRIPT

ferences between contours and PIs, not included.

For stops sharing the same place of articulation, the glottis is more constricted for voiced than for unvoiced cognates. Pharyngeal cavity, however, is larger in voiced when compared with unvoiced counterparts. For [p] the effect is observed through the entire pharynx, being for [t] and [k] differences more evident at oro-pharynx level.

The effect of coarticulation for stops is evident. For [k] the differences are more significant in the tongue tip region, since this articulator is free for the production of the vowel. For [t], the region with less variation is the one close to the place of articulation (dental), while tongue back is affected by the production of the vowel. In [p], the tongue is free for the production of the vowel, since [p] has a bilabial articulation.

5.3 Fricatives

The results for EP fricatives are presented in Fig. 16. Despite the non-inclusion of the superior incisors in the images, we can infer, through the position of the lips, that the [f] is produced through the approximation of lower lip to the upper incisors (labiodental fricative). Despite the fact that they are quite similar, our results point to an alveolar place of constriction for [s] and [z], being fricatives [ʃ] and [ʒ] produced slightly posterior. The differences for TT horizontal position between these fricatives are of only 6 mm, between [s] and [ʃ], and 4 mm for the other pair. The [s] production involves the tongue blade while, [ʃ] presents an apical articulation. Other differences between [s] and [ʃ] are: [s] is produced with a slightly lower TD position; the back of the tongue is more posterior in the production of [s]. The same pattern and articulation places can be observed for [z] and [ʒ]. These facts were confirmed using the superimposition of [s,ʃ] and [z,ʒ] midsagittal contours (not included). Through the analysis of the contours (not included) and their PIs, we observed that differences in configuration, for the same place of articulation and vocalic context, are not significant (in the midsagittal plane) in the unvoiced-voiced pairs. However, at the glottis level, there is a higher constriction for voiced fricatives, as already observed for voiced stops. Regarding pharyngeal cavity, there is a tendency for voiced fricatives to have a larger pharynx, but being the difference less evident than for stops.

We tested to see if our process was able to distinguish between the fricatives in three different VCV contexts, where V represents one of the vowels [a], [i], or [u]. The 2D results are presented in Fig. 17 and 3D results are shown in Fig. 18.

In Fig. 17, the effect of coarticulation is evident. In [f], a labiodental fricative,

ACCEPTED MANUSCRIPT

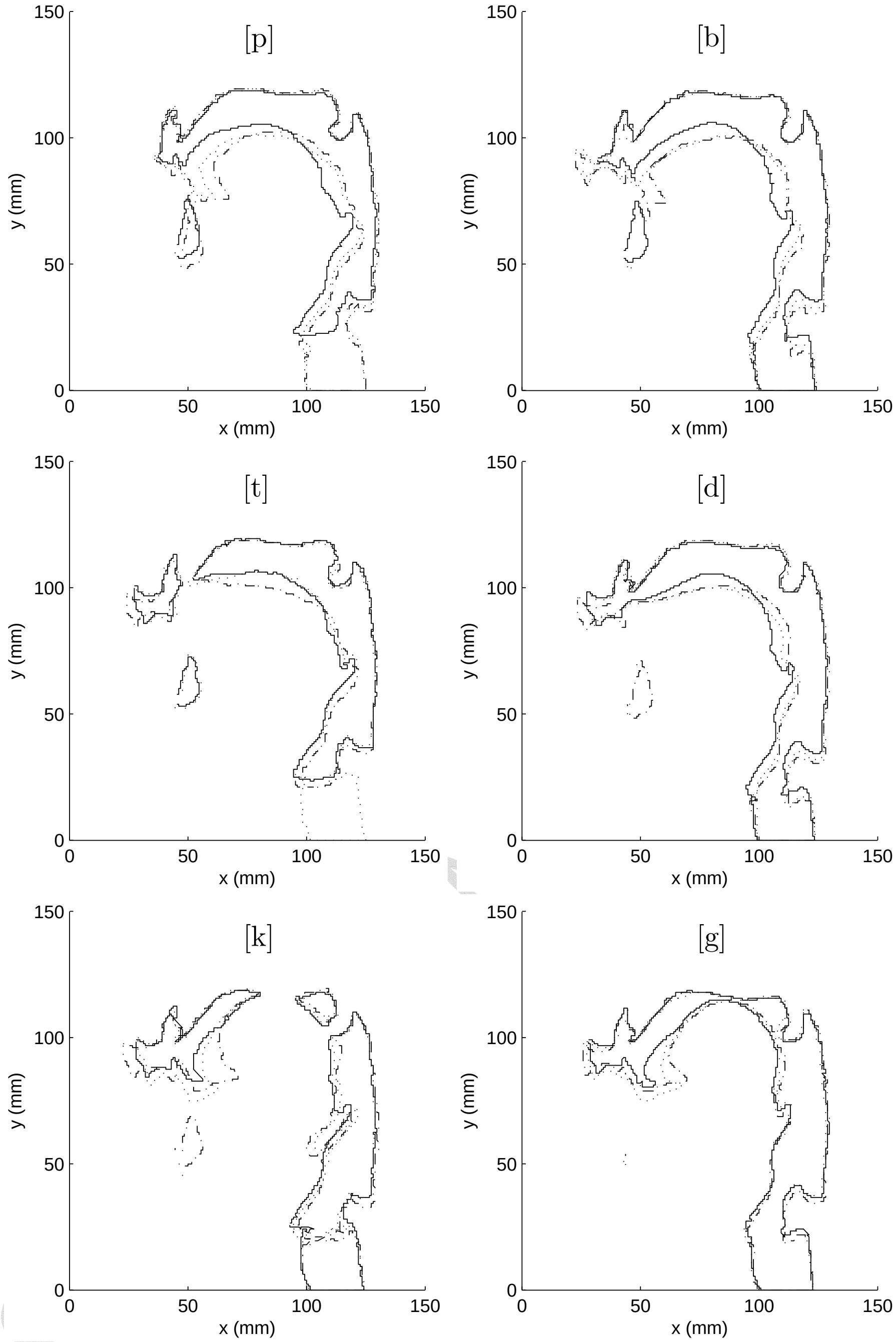


Figure 15. Midsagittal contours relative to stop consonants, obtained in VCV context with the point vowels [a] (dashed), [i] (solid line) and [u] (dash-dotted). At the top row appears the bilabial unvoiced [p] (left) and the voiced [b] (right); at center appear the dental unvoiced [t] (left) and the voiced stop [d] (right); at bottom the velar stops: the unvoiced [k] (left) and voiced [g] (right).

we observe differences both in tongue tip and tongue dorsum, the tongue being free for the production of the vowel. In [s], there are only differences in the posterior/back portion of the tongue. We do not observe the vowel effect on tongue tip or blade, used in the production of the consonant (apical alveolar).

ACCEPTED MANUSCRIPT

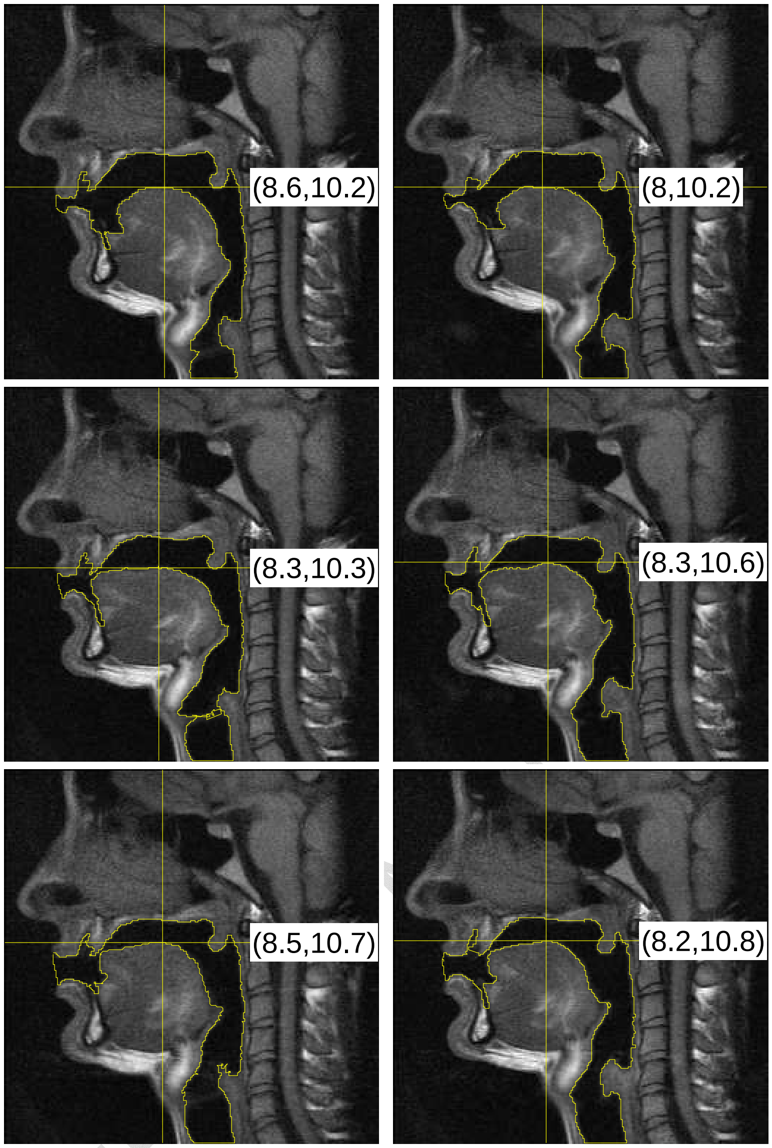


Figure 16. Midsagittal MRI images with superimposed contour relative to EP fricative sounds. At the top row the labiodental fricatives [f] and [v]; at the center the alveolar fricatives [s] and [z] and at bottom the palatoalveolar fricatives [ʃ] and [ʒ]. All were sustained having an example word with the fricative at the beginning and followed by the oral vowel [a].

Relative to [ʃ], the effect of the vowel in the tongue is even less visible. This sound, when compared with others in this study, presents a higher resistance to coarticulation.

For the voiced fricatives, the pattern of influence of the vowel in the production of the fricative consonant is similar to that observed for the unvoiced fricatives, being higher for the labiodental [v], smaller in the alveolar [z], and being [ʒ] production practically immune to the vowel effect.

ACCEPTED MANUSCRIPT

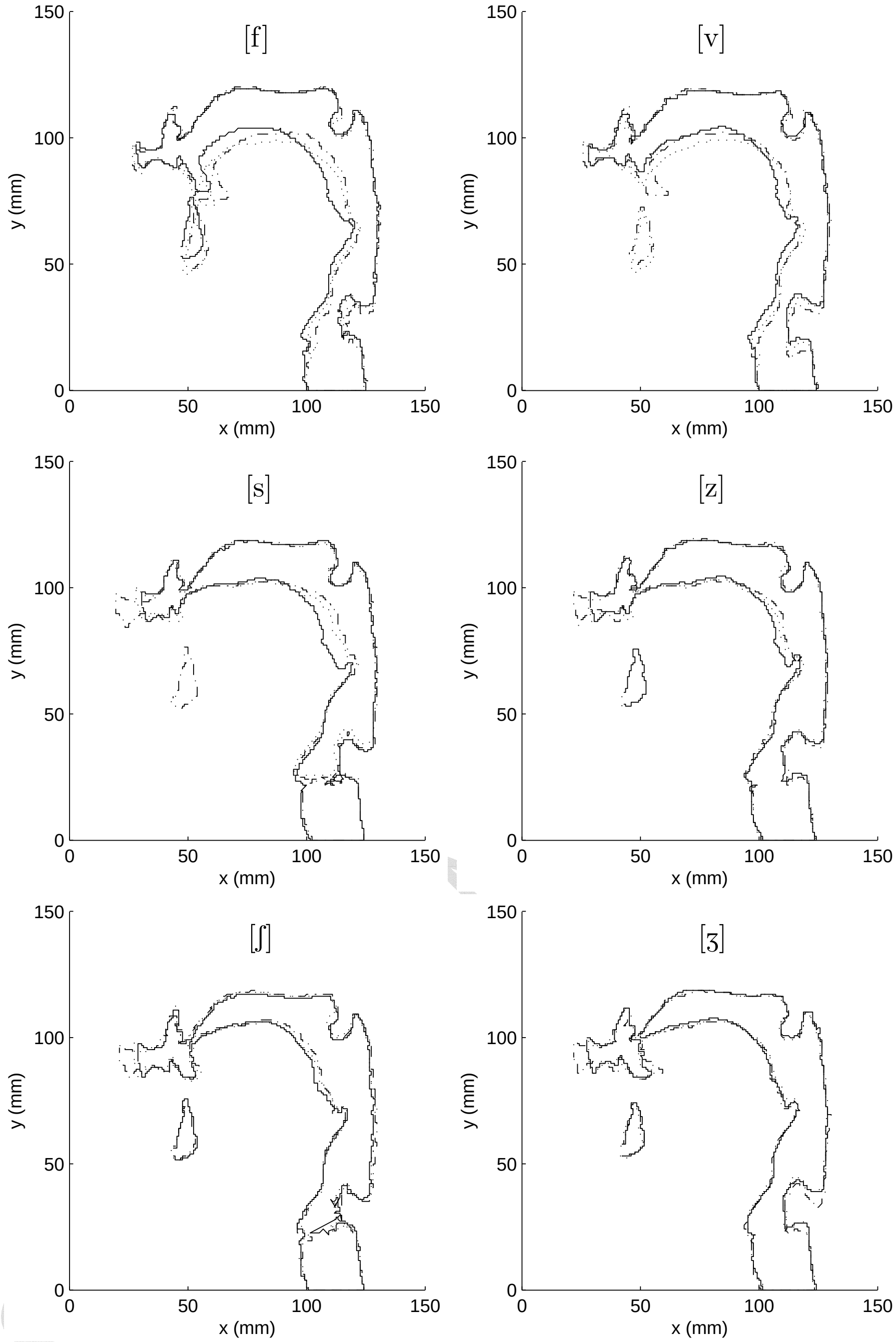
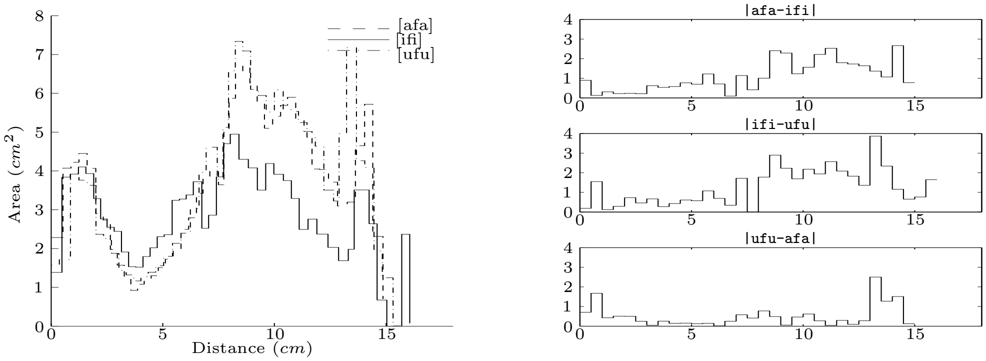


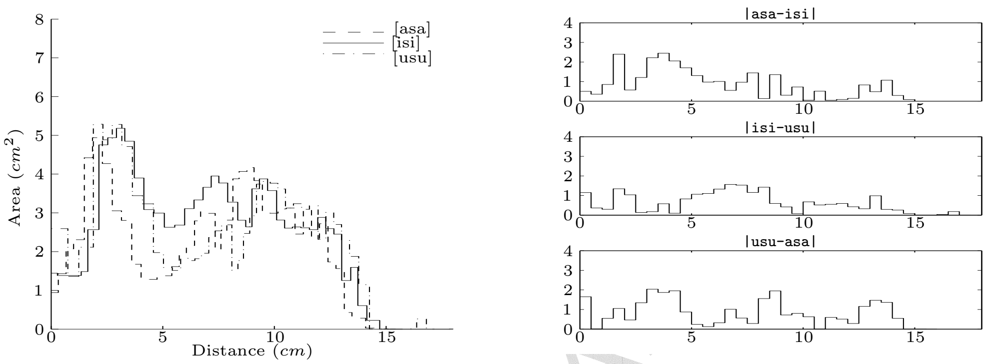
Figure 17. Midsagittal contours relative to fricatives, obtained in VCV context with the vowels [a] (dashed), [i] (solid line) and [u] (dash-dotted). At the top appears the labiodental unvoiced [f] (left) and the voiced [v] (right); in the middle row appear the alveolar unvoiced [s] (left) and alveolar voiced [z] (right); at bottom the palatoalveolar fricatives: the unvoiced [ʃ] (left) and the voiced [ʒ] (right).

Comparing the area functions and the differences between two area functions (average and maximum values), in Fig. 18, coarticulatory effects are smaller for [ʃ]. About the two other unvoiced fricatives, the most affected regions are

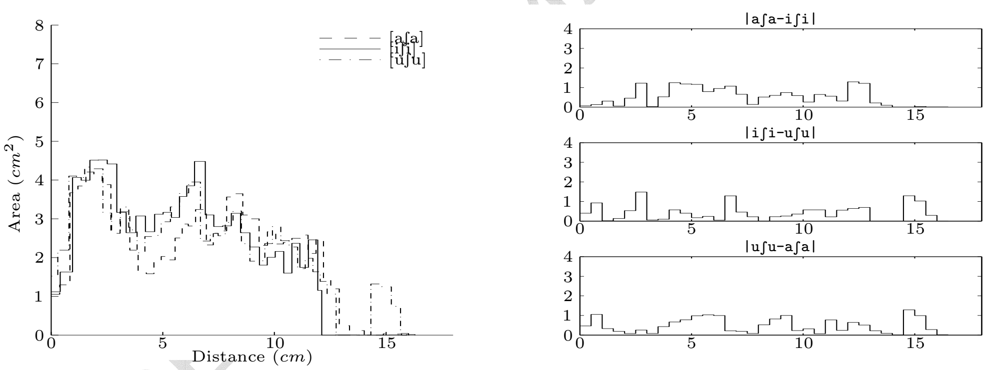
ACCEPTED MANUSCRIPT



(a) Fricative [f] in three vocalic contexts.



(b) Fricative [s] in three vocalic contexts.



(c) Fricative [ʃ] in three vocalic contexts.

Figure 18. Area functions for the fricatives [f], [s], and [ʃ] in three vocalic contexts (left) and absolute differences (right).

the pharyngeal region for [s] and the oral cavity for [f].

5.4 Laterals

The EP laterals, [l] and [ʎ], are shown in Fig. 19. Figure presents 2D information for [ʎ] and the two variants of the l-sound: [l] as in [lasu] and [ɫ] as in [maɫ]. For 3D, a third context is also included, intervocalic position ([palɐ]).

ACCEPTED MANUSCRIPT

In Fig. 20 we compare the 3 area functions obtained for [l].

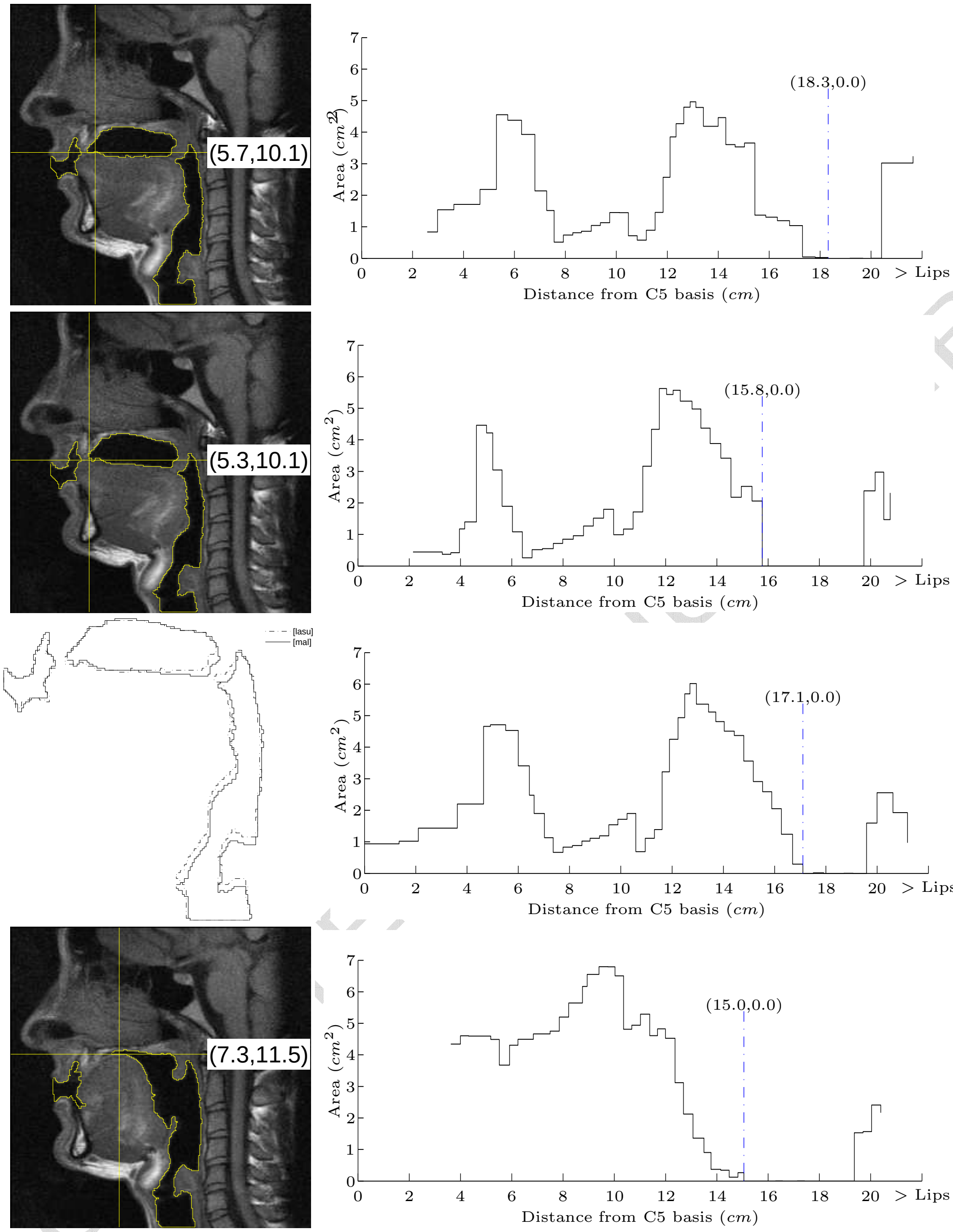


Figure 19. MRI images (with contours) and area functions for the EP laterals. Top 3 rows presents results for [l]: top row [l] in [lasu]; second row [ɫ] in [maɫ]; third row a comparison of the contours previously presented, on the left, and right, area function for a third context with only 3D data available, intervocalic position [palɐ]. Finally, on the bottom row, image and area function for [ʎ].

The first thing to note in Figs. 19 and 20 are the null areas in the area functions in the zone of partial occlusion. This is a result of the semiautomatic image processing, that was incapable of correctly segmenting the resliced images perpendicular to the centerline. Even with this limitation, 2D contours and

ACCEPTED MANUSCRIPT

area functions provide useful information on EP laterals.

Comparing the midsagittal profiles of the lateral [l] and [ɫ], we can verify that the place of articulation is the same for both sounds, in the alveolar/dental region. This can be confirmed both in contour superimposition and at the first point with null area in the area functions, all presented in Fig. 19. It is clear that the active articulator is tongue tip for both sounds.

Analyzing the area functions for [l] (Fig. 20), in the three contexts considered, we can observe a similar area variation pattern along the tract, without significant differences. We can report a constriction point beyond the lip region, corresponding to the alveolar area; upward in direction of the glottis an increase of area function is observed. This region corresponds to palatal area. A second constriction point is observed at uvular region, which is similar in the three positions. This second constriction is related with tongue dorsum raising. More detailed analyzes of tongue configurations on resliced coronal cuts, as in [7] and [64], are in progress.

The [ʎ] is usually described as a palatal consonant. When compared with the palatal [ɲ], [ʎ] has its occlusion more anterior. While in the area function the occlusion starts at 11.8 cm for [ɲ] (Fig. 13), for [ʎ] occlusion starts at 15.0 cm (Fig. 19). This points, at least for this speaker of EP, to a more palato-alveolar place of articulation for [ʎ]. It is produced with the tongue blade, the tongue dorsum not being in contact with the palate.

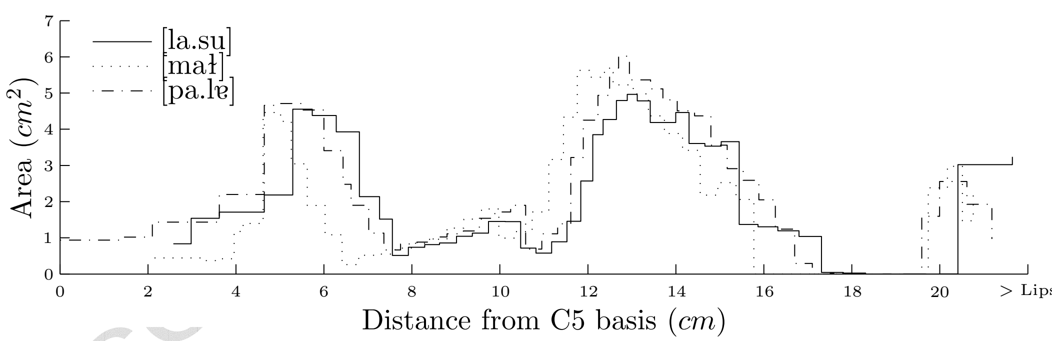


Figure 20. Comparison of the 3 area functions obtained for EP lateral [l]. Three contexts are represented: beginning of word and syllable (solid), end of word or syllable (dash-dotted) and in syllable onset but intervocalic (dotted).

6 Discussion

As our main objective is related to obtaining more data regarding EP production and not to exhaustively compare our results to published descriptions of EP, this discussion will not concentrate on pointing out all the agreements

ACCEPTED MANUSCRIPT

and disagreements between present work and EP common knowledge in articulatory phonetics. The availability of data for only one speaker also supports this option.

6.1 Corpus, MRI acquisition and Image processing

Our option to address as much as possible of EP sounds with only one speaker allowed us to cover, in a first study, what for other languages was produced incrementally. The existence of data regarding the several classes of EP sounds is particularly valuable to our work in articulatory synthesis. The disadvantage of only one speaker and the unique/reduced number of repetitions are, in our opinion, more than compensated by the advantages of the possibility of making direct comparison between different classes. This was particularly useful in the case of the comparative study of nasal vowels tract configuration relative to oral vowels; comparison of palatals [ɲ] and [ʎ] exact place of articulation and comparison of coarticulatory effects between stops and fricatives.

With our option for the (semi)automatic processing, the use of a direct 3D acquisition was possible. As the acquired MRI data are in a volumic layout, image processing techniques were necessary and sufficient means to create the appropriate reformatted planes for further segmentation. This additional flexibility makes it possible to obtain data in planes defined after acquisition and tuned to the objectives of the analyses. Moreover, there was a gain in the acquisition time. With this, our speaker had a much easier task and overall acquisition time was substantially reduced. The choice for a trained speaker with vocal and singing practice also contributed positively to a faster and less error prone acquisition. Some points need however improvement in the acquisition: improvement on the larynx region, sometimes affected by aliasing problems, to allow a better characterization of this zone of the oral tract; improve overall quality of the coronal images for a better study of laterals.

Semi-automatic image segmentation proved to be very useful and capable of attaining reproducible results. Neverthless, there are areas where improvements are needed: segmentation of the images in the zone of partial obstruction for laterals (not completely sucessful in this first approach); addition to the images of the separately acquired information on speakers' teeth.

6.2 Oral Vowels

One of the most relevant results obtained in this study, relative to EP oral vowels, is concerned with central vowels height. Contrary to traditional EP Phonetic descriptions (e.g. [73]), in which [ɨ] is considered as high as [i](anterior)

ACCEPTED MANUSCRIPT

and [u] (posterior) high vowels, we found that [ɨ] has, in fact, the highest TD position among the central vowels, but not so high to be considered a high central vowel. Only looking at jaw height (JH) alone we could describe [ɨ] as a closed vowel, similar to [i].

From an articulatory view point, the differences between the three central vowels are mainly related with tongue dorsum position and shape, jaw height and pharyngeal cavity dimensions (particularly the upper part). Amongst the three central vowels the one that is produced with the highest TD position is the [ɨ], followed by [ɐ] and [a]. Pharyngeal cavity dimension is also high for [ɨ] as the tongue dorsum is more raised and advanced in the production of this vowel, when compared with the other central vowels. Important characteristics of [a] are the very low jaw, high lip aperture and posterior position of tongue (TD and TR). The last characteristic goes against its classic classification of [a] as a central vowel, being better described as a low pharyngeal vowel. The [ɐ] is more similar to [a] than [ɨ] in terms of tongue shape; has an intermediate jaw opening, and presents lip aperture similar to [ɨ]. The [ɨ] appears as distinctively different from the other two vowels in the upper pharyngeal region, not presenting the characteristic narrowing of the others. These articulatory differences and characteristics of each of the 3 central vowels can be useful in clarifying their descriptions, a point of discussion in EP Phonetics. However, it is hard to generalize as our data are limited to one speaker. The dorsovelar location of the maximum constriction for the posterior vowel [o] is not in agreement with the usual articulatory description (e.g. [72]), reporting a pharyngeal location for the maximum constriction, as for [ɔ]. Obviously, due to corpus limitation to one speaker, we cannot clarify if this is a speaker characteristic, or a more general phenomenon.

6.3 Nasals

As expected, differences between nasal and oral vowels do not only concern velum lowering, but also differences in the position of other articulators [56]. The 2D results show that, at least with this speaker of EP, [ɐ̃] is markedly higher than [a]; [õ] is produced with an articulatory configuration between [o] and [ɔ]; and [ĩ] and [ũ] are produced with a height similar to the oral counterparts. These results agree in general with the ones obtained using EMMA and acoustic inference from first formant values [29]. When compared to French nasal vowels, some differences were detected, particularly at the pharyngeal cavity level. French nasal vowels seem to be produced with a more constricted pharyngeal region [13,51,56,57].

With the exception of [ɐ̃], a central vowel that presents the highest VPOQ, the posterior vowels ([ũ] and [õ]) have a slightly higher VPOQ than the anterior

ACCEPTED MANUSCRIPT

ones ([ĩ] and [ẽ]). The oral area is always higher than the nasal for all the sounds contemplated in our study, which implies a VPOQ smaller than 1. Although the VPOQ is smaller in orals, in our measures it was allways higher than zero due to the existence of a small passage to the nasal cavity even for the production of oral sounds. This is in agreement with the fact that nasal port opening is not sufficient to have a nasal sound. However, the VPOQ is an average value dependent of the sampling process, with possible failures in detecting nasal port closure. Comparing with recent results of Engwall, Delvaux and Metens [56], we verify that: the average VPOQ follows, in general terms, a similar behaviour: superior in nasal vowels than in the correspondent orals; the VPOQ values for French are significantly higher than the obtained for EP, particularly for the nasal vowels.

Relative to EP nasal consonants, the VPOQ results confirmed their relative position of velum aperture, between oral and nasal vowels. New 3D information contributed to validate previous work based on velum position only [28,29]. Also relevant is the close proximity of TD for [ɲ], [i] and [ĩ], consistent with the historic origin of the nasal consonant [ɲ].

6.4 Stops and Fricatives

Another fact that also deserves to be mentioned is related to the place of articulation of the so-called "velar" stop [k]. Contrary to the classical descriptions of [k], we observe that [k], at least for this speaker of EP, was produced in the palatal area and does not seem to be dependent on the vocalic context. Although the place of articulation of velar stops could vary with context [72], being more anterior when produced in the context of anterior vowels and more posterior in the context of back vowels, this is not observed in our study. In the different contexts studied, the place of articulation is always palatal, only with noticeable differences at tongue tip and blade level. In this area the effect of the vowel is clearly observed, the tip being more anterior in the context of [i] and more posterior in the context of [u]. Further studies are needed to clarify if this context independent point of constriction for [k] is (partially) related to the acquisition procedure, quite different from continuous speech.

For fricatives, [ʃ, ʒ] have the point of maximum constriction produced with the tongue tip slightly posterior relative to [s,z], but, in our opinion and using [32, p. 14] information on places of articulation, still in the alveolar region. This is not in accordance with what generally is described for [ʃ], as being produced by an approach of the tongue tip to the palato-alveolar or post-alveolar regions. A more detailed study of [ʃ] articulation point, using complementary techniques as EPG, should be considered.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Relative to the stridents, a great similarity in the place of articulation for [s, z] and for [ʃ,ʒ] was evident, the most obvious difference being at the level of sublaminal cavity which is larger for [ʃ] and [ʒ] than for [s] and [z]. This difference at the level of the sub-laminal cavity can be explained by the more apical articulation for [ʃ, ʒ], as the tongue tip is raised and slighly more posterior. These results are only partially in line with previous results reported for fricatives, but for a different language [49].The authors reported for [ʃ,ʒ] a high tendency for a laminal articulation rather than apical, and referred to a speaker dependent variability for [s,z] with respect to apical and laminal articulations.

Our results regarding a more constricted glottis region together with a larger pharynx for voiced sounds are in line with what was reported by Narayanan et al [49], for fricatives: a tendency for larger pharyngeal areas for voiced sounds. This fact was also previously reported by Perkell [74] for the sibilants [s] and [z] using X-ray techniques. This constriction at glottis level together with a larger pharynx might be explained by the necessity of having muscular adjustments and adequate pressure differences to produce phonation in voiced sounds.

6.5 Laterals

In laterals, the differences between [l] and [ɫ] are not significant considering both 2D and 3D information. For American English, as reported by Narayanan et al. [7] and Bangayan et al. [64], there are differences in the back region for light and dark versions. For EP, we found /l/ velarization not only in syllable final position, as expected, but also in syllable initial position. EP area functions (for all the contexts considered for /l/) present a similar pattern in front and back regions, which means a second constriction point independent of position in the syllable (onset or coda). These facts point to the existence of only one positional allophone for /l/, a dark , which is in line with Andrade [33] descriptions for EP: velarization occurs not only in syllable final position but also in initial position. This is also in agreement with older descriptions, see Strevens [26] and section 1.2.

As far as [ʎ] is concerned, our results point to a more anterior place of articulation (alveolopalatal) instead of palatal, which is not in line with EP most frequent descriptions, already referred to in the introduction. However, Sá Nogueira [75] has already pointed to the possibility of this consonant having a more anterior place of articulation. Our finding is also in agreement with what was reported by Recasens and Espinosa (2006) [76]. These authors referred the fact that the lateral [ʎ] cannot be exclusively articulated in the palatal area. They pointed out that Romanic Languages also present a closure in the alveolo-palatal area, that could even be alveolar. When compared with the

ACCEPTED MANUSCRIPT

palatal [ɲ], it is evident a more anterior articulation point for [ʎ] and a "closure fronting decreasing in the progression [ʎ] > [ɲ]" as also reported by these authors [76].

6.6 Coarticulation

In general, EP stops are less resistant to coarticulatory effects than EP fricative. This is in agreement with the less constrained tongue body for stops, when compared to fricatives, reported for other languages by Farnetani [77] and Recasens [78]. Comparing the labiodental fricative [f] with the bilabial stop [p] it is observed that the effect of the adjacent vowel is greater on the stop than on the fricative of the corresponding class. However, this difference is still sharper when we compare, by e.g., the alveolar fricative [z] with the dental stop [t].

In our study, concerning the tongue blade, for the stops [t,d] and the fricative [s] there is no significant effect of the vowel in this region, although the influence is evident in the production of the stops [k] and [g]. Recasens [78] reports that the tongue region can present different articulatory behavior as a function of its evolvement in the production of a certain configuration. It is predicted that the blade must be more resistant to coarticulation during the production of alveolar consonants [t,d,s] than on the velar [k,g]. This is also verified in our study.

Among all the sounds studied here, and not considering any articulator in particular, the sounds that have the highest resistance to coarticulation are [ʃ] and [ʒ]. This fact was already observed by Farnetani [77] and can be connected with the complexity involved in the production of these sounds, [79]. Recasens et al. [38] also refers to the fact that some sounds are more constrained than others.

In accordance with Kiritani [80], we can also consider the tongue-jaw system together. We verified that velar consonants [k] and [g] in [i] context present a more anterior position of tongue blade, but this anteriorization is not evident at jaw level. Tuller et al. [81], also stated that the height of the jaw does not change in VCV context for [t] and [f], but suffers alterations due to the vowel in [p] and [k]. In our corpus, it was verified that for [t] there is no alteration in the height of the jaw, but this is seen in the production of [f].

ACCEPTED MANUSCRIPT

7 Conclusions

In this paper we present new MRI data relative to the majority of the EP sounds. Both 2D and 3D MRI data are provided. In line with other studies in the field for other languages, we obtained volumetric MRI but using a different and faster acquisition technique. Unlike other studies in this field, we have used a semiautomatic segmentation method.

MRI data obtained for one EP speaker, complemented by the utilization of imaging processing techniques and analyses, was determinant to improve our knowledge on EP oral and nasal sounds, laterals, fricatives and stops. With 2D MRI data, we compared oral and nasal vowels contours, leading to more detailed information than previously possible with other techniques such as EMMA. 3D information and area functions revealed very useful for palatal sounds [ɲ] and [ʎ], characteristic of EP. This is valuable information for evolution of articulatory synthesis of European Portuguese. Also, without claiming generalization due to the single speaker limitation of the data, some interesting findings were reported for palatal consonants, central vowels and laterals. It was possible to verify, for the EP, some facts related to coarticulation already reported for other languages. These results are also interesting due to the reduced use of MRI in coarticulation studies.

7.1 Future

With this study, the capacities of MRI in providing useful information on speech production, particularly for EP or in general, is far from being exhausted. After this broad study, we consider as important the following possible continuations:

- Perform a formal evaluation of 3D segmentation method, not yet performed due to time limitations;
- Improve the area function computation regarding speed, accuracy in the laryngeal region, and taking in consideration the teeth. Only with an improved acquisition and segmentation of the tract near the larynx will be possible to solve the current limitations on area functions length and origin;
- Process the nasal tract 3D acquisition to obtain nasal tract area function;
- Complement the comparisons between nasal and oral vowels with realtime MRI information. Despite usefull for the characterization of EP nasal vowels, the information available for this study suffers from two important limitations: only one speaker was recorded and the variation over time of vocal tract is not available. Realtime MRI, with adequate time resolution and from several speakers, is needed to reduce the remaining doubts regarding

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

the nasal vowels tract configuration;

- Conduct specific studies addressing a sound class or set of sounds in detail, with several repetitions and a reasonable number of speakers. This can be started by studying the EP laterals for which we had interesting results, needing more data to enable any generalization;
- Repeat acquisition of the present corpus with more speakers. This is necessary to solve the speaker dependent nature of the reported results. Provision to include speakers from different dialects should be considered. With information regarding several speakers and the associated contours and area functions, a search for representative shape descriptors should be investigated;
- Complement the study using real time MRI. Real time acquisition with a corpus mainly composed of nasal sounds and trills has already been carried out, but not yet fully analysed. In this preliminary and first approach we obtained a temporal resolution close to 200 ms (5 frames/s). We are particularly interested in improving temporal resolution and obtaining dynamic information on articulators movements, particularly for nasals, during actual production of EP words. Coarticulatory effects will greatly benefit from this line of research.

8 Acknowledgements

This work is part of project HERON (POSI/PLP/57680/2004), funded by FCT (Portuguese Research Agency). Authors thank Radiology Department, Coimbra University Hospital (HUC), particularly its Director Professor Filipe Caseiro Alves and its technical staff. We gratefully acknowledge the very important MRI technical support given by João Cunha Pires. We also thank our two speakers for their help and tolerance during the acquisition session. Our thanks to the 3 anonymous reviewers for their comments and suggestions, contributing to an overall improvement in the paper.

References

[1] A. Teixeira, R. Martinez, L. N. Silva, L. M. T. Jesus, J. C. Príncipe, F. Vaz, Simulation of human speech production applied to the study and synthesis of European Portuguese, EURASIP Journal on Applied Signal Processing 2005 (9) (2005) 1435–1448.

[2] P. Hoole, Methodological considerations in the use of electromagnetic articulography in phonetic research, FIPKM 31 (1993) 43–64.

ACCEPTED MANUSCRIPT

[3] P. Hoole, N. Nguyen, Electromagnetic articulography, in: W. Hardcastle, N. Hewlett (Eds.), Coarticulation: Theory, Data and Techniques, Cambridge University Press, Cambridge, 1999, pp. 260–269.

[4] M. Stone, Laboratory techniques for investigating speech articulation, in: W. H. Laver, John (Eds.), The Handbook of Phonetic Sciences, Blackwell, 1999, pp. 11–32.

[5] T. Baer, J. C. Gore, L. C. Gracco, P. W. Nye, Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels, Journal of the Acoustical Society of America (JASA) 90 (2) (1991) 799–828.

[6] A. Alwan, S. Narayanan, K. Haker, Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics, Journal of the Acoustical Society of America (JASA) 101 (2) (1997) 1078–1089.

[7] S. Narayanan, A. Alwan, K. Haker, Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals, Journal of the Acoustical Society of America (JASA) 101 (2) (1997) 1064–1077.

[8] S. Narayanan, K. Nayak, S. Lee, D. Byrd, An approach to real-time magnetic resonance imaging for speech production, Journal of the Acoustical Society of America (JASA) 115 (4) (2004) 1771–1776.

[9] M. Tiede, S. Masaki, E. Vatikiotis-Bateson, Contrasts in speech articulation observed in sitting and supine conditions, in: 5th Seminar on Speech Production, Kloster Seeon, Germany, 2000, pp. 25–28.

[10] O. Engwall, A revisit to the application of MRI to the analysis of speech production - testing our assumptions, 6th Seminar on Speech Production (2003) 43–48.

[11] S. Narayanan, A. Alwan, A nonlinear dynamical systems analysis of fricative consonants, Journal of the Acoustical Society of America 97 (1995) 2511–2524.

[12] M. Stone, A. J. Lundberg, E. P. Davis, R. Gullipalli, M. NessAiver, Three-dimensional coarticulatory strategies of tongue movement, in: 5th European Conference on Speech Communication and Technology (Eurospeech), 1997, pp. 1–31.

[13] D. Demolin, M. George, V. Lecuit, T. Metens, A. Socquet, Détermination, par IRM, de l'ouverture au velum des voyelles nasales du Français, in: XXIèmes Journées d'Etudes sur la Parole, Avignon, France, 1996, pp. 83–86.

[14] P. Badin, G. Bailly, M. Raybaudi, C. Segebarth, A three-dimensional linear articulatory model based on MRI data, in: 5th International Conference on Spoken Language Processing (ICSLP), 1998, pp. 417–420.

[15] A. Serrurier, P. Badin, Towards a 3D articulatory model of the velum based on MRI and CT images, ZAS Papers in Linguistics 40 (2005) 195–211.

ACCEPTED MANUSCRIPT

[16] O. Engwall, P. Badin, Collecting and analysing two and three dimensional MRI data for Swedish, Speech Transmission Laboratory: quarterly Progress and Status Report (STL-QPSR) (1999) 11–38.

[17] C. Ericsdotter, Articulatory-acoustic relationships in Swedish vowel sounds, Doctoral dissertation, Stockholm University (2005).

[18] H. Takemoto, T. Kitamura, H. Nishimoto, K. Honda, A method of tooth superimposition of MRI data for accurate measurement of vocal tract shape and dimensions, Acoustical Science and Technology 25 (6) (2004) 468–474.

[19] B. J. Kröger, R. Winkler, C. Mooshammer, B. Pompino-Marschall, Estimation of vocal tract area function from magnetic resonance imaging: preliminary results, in: 5th Seminar on Speech Production, Kloster Seeon, Germany, 2000, pp. 333–336.

[20] P. Hoole, A. Wismüller, G. Leinsinger, C. Kroos, A. Geumann, M. Inoue, Analysis of tongue configuration in multi-speaker, multi-volume MRI data, in: 5th Seminar on Speech Production, Kloster Seeon, Germany, 2000, pp. 157–160.

[21] K. Mathiak, I. Hertrich, W. E. Kincses, U. Klose, H. Ackermann, W. Grodd, Stroboscopic articulography using fast magnetic resonance imaging, International journal of language & communication disorders/Royal College of Speech & Language Therapists 35 (3) (2000) 419–25.

[22] M. Tiede, An MRI-based study of pharyngeal volume contrasts in Akan and English, Journal of Phonetics 24 (4) (1996) 399–421.

[23] A. Teixeira, F. Vaz, European Portuguese nasal vowels: an EMMA study, in: 7th European Conference on Speech Communication and Technology (EuroSpeech), Vol. 2, Scandinavia, 2001, pp. 1483–1486.

[24] S. Rua, D. Freitas, Morphological dynamic study of human vocal tract, in: CompIMAGE - Computational Modelling of Objects Represented in Images: fundamentals, Methods and Applications, Coimbra, Portugal, 2006.

[25] F. N. Gregio, Configuração do trato vocal supraglótico na produção das vogais do português brasileiro: dados de imagens de ressonância magnética, Dissertação de mestrado, Pontificia Universidade Católica de São Paulo (2006).

[26] P. Strevens, Some observations on the phonetics and pronunciation of modern Portuguese, Revista do Laboratório de Fonética Experimental de Coimbra II (1954) 5–29.

[27] M. Cruz-Ferreira, Portuguese (European), in: Handbook of the International Phonetic Association, The International Phonetic Association, Cambridge University Press, 1999, pp. 126–130.

[28] S. Rossato, A. Teixeira, L. Ferreira, Les nasales du portugais et du français: une étude comparative sur les données EMMA, in: XXVIes Journées d´études sur la parole, Dinard, 2006.

ACCEPTED MANUSCRIPT

[29] A. Teixeira, L. Castro Moutinho, R. L. Coimbra, Production, acoustic and perceptual studies on European Portuguese nasal vowels height, in: Int. Congress Phonetic Sciences (ICPhS), 2003, pp. 3033–3036.

[30] A. Teixeira, F. Vaz, J. C. Príncipe, Influence of dynamics in the perceived naturalness of Portuguese nasal vowels, ICPhS (1999) 2557–2560.

[31] L. M. T. Jesus, C. H. Shadle, A parametric study of the spectral characteristics of European Portuguese fricatives, Journal of Phonetics 30 (2002) 437–464.

[32] P. Ladefoged, I. Maddieson, The Sounds of the World's Languages, Blackwell, 1996.

[33] A. Andrade, On /l/ velarization in European Portuguese, in: International Conference of Phonetics (ICPhS), San Francisco, 1999.

[34] D. Recasens, A. Espinosa, Articulatory, positional and coarticulatory characteristics for clear /l/ and dark /l/: Evidence from two catalan dialects, Journal of the International Phonetic Association 35 (1) (2005) 1–25.

[35] B. Kühnert, F. Nolan, The origins of coarticulation, in: W. H. Hewlett, Nigel (Eds.), Coarticulation: theory, data and techniques, Cambridge University Press, 1999.

[36] H. S. Magen, The extent of vowel-to-vowel coarticulation in english, Journal of Phonetics 25 (2) (1997) 187–205.

[37] S. Manuel, Cross-language studies: relating language-particular coarticulation patterns to other language-particular facts, in: W. J. H. Hewlett, Nigel (Eds.), Coarticulation: theory, data and techniques, Cambridge University Press, Cambridge, 1999, Ch. 8, pp. 179–198.

[38] D. Recasens, M. D. Pallarès, J. Fontdevila, A model of lingual coarticulation based on articulatory constraints, Journal of the Acoustical Society of America (JASA) 102 (1997) 544–561.

[39] W. Hardcastle, N. Hewlett, Coarticulation: theory, data and tecniques, Cambridge University Press, Cambridge, 1999.

[40] P. West, Long-distance coarticulatory effects of British English /l/ and /r/: an EMA, EPG and acoustic study, in: Speech Production Seminar, Seeon, Germany, 2000, pp. 105–108.

[41] O. Engwall, P. Badin, An MRI study of Swedish fricatives: coarticulatory effects, in: 5th Seminar on Speech Production, Kloster Seeon, Germany, 2000, pp. 297–300.

[42] M. Stone, E. P. Davis, A. S. Douglas, M. NessAiver, R. Gullipalli, W. S. Levine, A. J. Lundberg, Modeling tongue surface contours from Cine-MRI images, Journal of Speech, Language, and Hearing Research 44 (2001) 1026–1040.

[43] O. Engwall, Tongue talking: studies in intraoral speech synthesis, Doctoral thesis, KTH - Royal Institute of Technology (2002).

ACCEPTED MANUSCRIPT

[44] B. H. Story, I. R. Titze, E. A. Hoffman, Vocal tract area functions from magnetic resonance imaging, Journal of the Acoustical Society of America (JASA) 100 (1) (1996) 537–554.

[45] A. R. Greenwood, C. C. Goodyear, P. A. Martin, Measurements of vocal tract shapes using magnetic resonance imaging, in: IEE: Communications Speech & Vision, Vol. 139, 1992, pp. 553–560.

[46] B. Yang, Measurement and synthesis of the vocal tract of Korean monophthongs by MRI, XIVth International Congress of Phonetic Sciences (ICPhS) (1999) 2005–2008.

[47] C. H. Shadle, M. Mohammad, J. N. Carter, P. J. B. Jackson, Multi-planar dynamic magnetic resonance imaging: new tools for speech research, XIVth International Congress of Phonetic Sciences (ICPhS) (1999) 623–626.

[48] A. Serrurier, P. Badin, A three-dimensional linear articulatory model of velum based on MRI data, in: Interspeech, 2005.

[49] S. Narayanan, A. Alwan, K. Haker, An articulatory study of fricative consonants using MRI, Journal of the Acoustical Society of America (JASA) 98 (3) (1995) 1325–1347.

[50] D. Demolin, T. Metens, A. Soquet, Three-dimensional measurement of the vocal tract by MRI, in: 4th International Conference on Spoken Language Processing (ICSLP), Vol. 1, 1996, p. 272.

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

[52] J. Dang, K. Honda, MRI measurements and acoustic of the nasal and paranasal cavities, Journal of the Acoustical Society of America (JASA) 94 (3, Pt 2) (1994) 1765.

[53] J. Dang, K. Honda, An improved vocal tract model of vowel production implementing piriform resonance and transvelar nasal coupling, in: ICSLP, 1996.

[54] O. Engwall, Modeling of the vocal tract in three dimensions, in: 6th European Conference on Speech Communication and Technology (Eurospeech), 1999, pp. 113–116.

[55] D. Demolin, V. Lecuit, T. Metens, B. Nazarian, A. Soquet, Magnetic resonance measurements of the velum port opening, in: 5th International Conference on Spoken Language Processing (ICSLP), 1998.

[56] O. Engwall, V. Delvaux, T. Metens, Interspeaker variation in the articulation of nasal vowels, in: 7th International Seminar on Speech Production, 2006.

[57] V. Delvaux, T. Metens, A. Soquet, French nasal vowels: acoustic and articulatory properties, in: 7th International Conference on Spoken Language Processing (ICSLP), Vol. 1, Denver, 2002, pp. 53–56.

ACCEPTED MANUSCRIPT

[58] J. Dang, K. Honda, H. Suzuki, Morphological and acoustical analysis of the nasal and the paranasal cavities, Journal of the Acoustical Society of America (JASA) 96 (4) (1994) 2088–2100.

[59] H. Kim, Stroboscopic-cine MRI data on Korean coronal plosives and affricates: implications for their place of articulation as alveolar, Phonetica 61 (4) (2004) 234–251.

[60] C. H. Shadle, M. Tiede, S. Masaki, Y. Shimada, I. Fujimoto, An MRI study of the effects of vowel context on fricatives, in: Institute of Acoustics, Vol. 18, 1996, pp. 187–194.

[61] M. Mohammad, E. Moore, J. N. Carter, C. H. Shadle, S. R. Gunn, Using MRI to image the moving vocal tract during speech, in: 5th European Conference on Speech Communication and Technology (Eurospeech), Vol. 4, 1997, pp. 2027–2030.

[62] P. J. B. Jackson, Characterisation of plosive, fricative and aspiration components in speech production, Ph.D. thesis, U. Southampton (2000).

[63] S. Narayanan, A. Alwan, Noise source models for fricative consonants, IEEE Transactions on Speech and Audio Processing 8 (2) (2000) 328–344.

[64] P. Bangayan, A. Alwan, S. Narayanan, From MRI and acoustic data to articulatory synthesis: a case study of the laterals, in: ICSLP, Philadelphia, 1996, pp. 793–796.

[65] B. Gick, A. M. Kang, D. H. Whalen, MRI evidence for commonality in the post-oral articulations of English vowels and liquids, Journal of Phonetics 30 (3) (2002) 357–371.

[66] International Phonetic Association, Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet, Cambdridge University Press, 1999.

[67] B. Santos, C. Ferreira, J. Silva, A. Silva, L. Teixeira, Quantitative evaluation of a pulmonary contour segmentation algorithm in x-ray computed tomography images, Academic Radiology 11 (8) (2004) 868–878.

[68] A. Chodorowski, U. Mattsson, M. Langille, G. Hamarneh, Color lesion boundary detection using live wire, in: SPIE, 2005.
URL `http://www.cs.sfu.ca/~hamarneh/software/livewire/index.%html`

[69] R. Adams, L. Bischof, Seeded region growing, IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (6) (1994) 641–647.

[70] L. Sachs, Applied Statistics - A Handbook of Techniques, 2nd Edition, Springer-Verlag, 1984.

[71] A. Bryman, D. Cramer, Quantitative Data Analysis with SPSS Release 10 for Windows - A guide for Social Scientists, Routledge, 2001.

ACCEPTED MANUSCRIPT

[72] A. Morais Barbosa, Introdução ao Estudo da Fonologia e Morfologia do Português, Almedina, Coimbra, 1994.

[73] M. d. C. Viana, A. Andrade, Fonética, in: I. Faria, E. Pedro, I. Duarte, C. Gouveia (Eds.), Introdução à Linguística Geral e Portuguesa, Caminho, Lisbon, 1996, pp. 113–167.

[74] J. Perkell, Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study, MIT Press, 1969.

[75] R. Sá Nogueira, Elementos para um tratado de Fonética Portuguesa, Impressa Nacional, Lisbon, 1938.

[76] D. Recasens, A. Espinosa, Articulatory, positional and contextual characteristics of palatal consonants: Evidence from Majorcan Catalan, Journal of Phonetics 34.

[77] E. Farnetani, Coarticulation and connected speech processes, in: W. J. H. Laver, John (Eds.), The Handbook of Phonetic Sciences, Blackwell, Oxford, 1999, pp. 371–404.

[78] D. Recasens, Lingual coarticulation, in: W. Hardcastle, N. Hewlett (Eds.), Coarticulation, Cambridge University Press, Cambridge, 1999.

[79] W. J. Hardcastle, Physiology of Speech Production: an Introduction for Speech Scientists, Academic Press, London, 1976.

[80] S. Kiritani, X-ray microbeam method for measurement of articulatory dynamics-techniques and results, Speech Communication 5 (2) (1986) 119–140.

[81] E. Tuller, K. S. Harris, R. Gross, Electromyographic study of the jaw muscles during speech, Journal of Phonetics 9 (1981) 175–188.

ACCEPTED MANUSCRIPT