An MRI study of European Portuguese nasals

Paula Martins¹, Inês Carbone², Augusto Silva², António Teixeira²

¹Escola Superior de Saúde, Universidade de Aveiro, Portugal
²DEI/IEETA, Universidade de Aveiro, Portugal

pmartins@essua.ua.pt, inesdomingues@gmail.com, asilva@det.ua.pt, ajst@det.ua.pt

Abstract

In this work we present a recently acquired MRI database for European Portuguese. As a first example of possible studies, we present results on 2D and 3D analyses of European Portuguese nasals, particularly nasal vowels. This database will enable the extraction of 2D and/or 3D articulatory parameters as well as some dynamic information to include in articulatory synthesizers. It can also be useful to compare the production of European Portuguese with the production of other languages and have further insight on some of the European Portuguese characteristics, as the nasalization and coarticulation. The MRI database and related studies were made possible by the interdisciplinary nature of the research team, comprised of a radiologist, image processing specialists and a speech scientist.

Index Terms: MRI, European Portuguese, Nasals

1. Introduction

Building phonetic information databases has great relevance in fields such as speech synthesis, speech recognition, speech disorders studies, learning of new languages, etc. We have been involved in articulatory synthesis for over a decade, and it is an area where data production is very important [1]. This type of anthropomorphic synthesizers demands large amounts of detailed anatomic-physiological information, if possible in 3D, and their variation in time. For European Portuguese (EP), not much information is available.

Nowadays, the most common methods found in speech research literature to acquire information are: Electromagnetic Articulography (EMMA), Electropalatography (EPG) and Magnetic Resonance Imaging (MRI). For EP, one of the authors was involved in the creation of an EMMA database, focused on nasals [2]. Despite being a very complete database, EMMA acquisition method has some drawbacks: measurements are generally limited to two dimensions and data is point-wise [2]. EPG measures only the linguopalatal contact with its variation on time, being difficult to make well-fitted pseudo-palates which in turn interfere to some extent with speech production. Also, to the best of our knowledge, there are no EPG databases for EP. MRI, the technique on which we focus our study, has some potential advantages: good contrast between soft tissues, 3D modeling and capacity to observe the vocal tract in all of its extension. This last advantage is of special interest in the study of the pharyngeal cavity, since it is not accessible through EMMA or EPG. Moreover, it is a non-invasive and relatively safety image technique. Disadvantages are related to the absence of the teeth in the images, due to their lack of Hydrogen protons, and the acquisition technique, in which the speaker must be lying down during the speech production. This position can have some influence, for instance, on the tongue posture, [3], but this drawback can be considered acceptable. The relatively low temporal resolution achieved, even with the fastest techniques, is also a limiting factor.

Nasality is present in most (97%) world languages, even thought only about 20% of such languages have nasal vowels [4]. Portuguese and French languages present a nasalization both on vowel and consonantal systems [4]. EP has five nasal vowels, [i], [e], [ɛ], [u], [i]: three nasal consonants, [n], [ŋ], [ɲ]; and several nasal diphthongs and triphthongs. However, some phonetic studies point to the existence of differences related with production mechanisms. This special characteristic motivated the research on the EP nasal sounds by Teixeira [5, 2]. There is some uncertainty as to 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 [6]. In this work, we return to the same challenging topic, using MRI as a data acquisition method.

EP nasals have been acoustically studied in the past and recently, as mentioned before, using EMMA [2]. No use of MRI in EP nasals studies was reported to our knowledge. For French nasal sounds, several studies were already conducted, focused on the velum opening, measurements on pharyngeal cavity for nasal vowels [7], 3D articulatory models of velum [8], determination of Velum Port Opening Quotient (VPOQ), and nasal tract areas amongst different speakers [9].

In this work, we describe the MRI corpus acquired for the EP and show some results related with nasal sounds. Section 2 includes specifications of the corpus, with reference to its main objectives, contents, acquisition method, and the informants. Section 3 explains the image processing techniques and in Section 4 we show some preliminary results derived from 2D and 3D corpus. In Sections 5 and 6 we finalize with the discussion, conclusions and future work.

2. Corpus

2.1. Design/Objectives

The main objective of the corpus was the acquisition of anatomic, physiological and dynamic information related with the production of the EP sounds. As specific objectives, we would like to obtain contours, articulatory parameters, area functions, and quantification of the velum port opening and 2D/3D models of the vocal tract.

2.2. Contents

The corpus is comprised of three parts: 2D acquisition in the sagittal plane (A), 3D acquisitions (B) and acquisition in real-time (C). In the static corpus (A and B), the sounds are arti-
ficially sustained or holding the articulation during the period of image acquisition, as already done in a similar way for other languages [10]. In this paper we present and use only the part of the corpus relative to nasal sounds and oral vowels. Acquisition also contemplated EP consonants (stops, fricatives and laterals).

**2D corpus:** Each sound of the 2D and 3D corpus was produced in the context of a reference word The description of the 2D corpus can be found in Table 1. Each phone was repeated only once.

<table>
<thead>
<tr>
<th>Phoneme</th>
<th>2D Corpus Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>[a] [i] [r] [x]</td>
</tr>
<tr>
<td>Nasals</td>
<td>[i] [e] [o]</td>
</tr>
<tr>
<td>C</td>
<td>[n] [n] [p]</td>
</tr>
</tbody>
</table>

**3D Corpus:** In this part we opted for a different approach than the one found in the literature, e. g. [10]. Instead of choosing a set of directions and acquiring a fixed number of slices, we have used a 3D sequence (see Section 2.3 for more information on the acquisition process). The main advantage is that with this technique we can reslice the volume in any direction. More details on this corpus can be found in Table 2. As previously, each phone was repeated only once.

<table>
<thead>
<tr>
<th>Phoneme</th>
<th>3D Corpus Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>[a] [u] [x]</td>
</tr>
<tr>
<td>Nasals</td>
<td>[i] [e] [o]</td>
</tr>
<tr>
<td>C</td>
<td>[n] [n] [p]</td>
</tr>
</tbody>
</table>

**Real-time corpus:** With this corpus we intended to study dynamic productions, acquired in real time. The words chosen were selected from the EMMA corpus [2], consisting of word pairs differing only in the vowel (ex: [k̪atu]/[katu] and [m̪atu]/[matu]). We also included some words enabling the study of the EP rothics (ex: [kanu]/[kaRu]).

For nasal sounds, we also acquired images with an oblique cut passing thought the velum. With these sequences, we can obtain dynamic temporal information of the opening and closing movement of the velum. In this paper, we will not show these results.

### 2.3. MRI acquisition

The images were acquired using a 1.5 Tesla (Magnetom Symphony, Maestro Class, Siemens, Erlanger, Germany) scanner equipped with Quantum 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.

The 2D corpus was acquired in the sagittal plane with a TSE T1 weighted sequence (Slice thickness = 5 mm, FOV = 240 mm, Matrix = 192x256, ETL = 15). Each of these images took about 5.6 seconds to be acquired.

Instead of using a multislice imaging method, obtaining a stack of two dimensional images, to extract 3D information, we decided for a true three-dimensional image acquisition method (3D FT). This method uses a broadband, non-selective RF pulse to excite a volume of spins simultaneously. This 3D acquisition method allows [11]: (1) reconstruction of very thin slices with good detail in any direction (useful to obtain slices orthogonal to vocal tract centerline); (2) higher SNR when compared with 2D imaging; (3) faster and easier 3D visualization is possible.

Based on the above considerations, the 3D corpus was acquired using a fast spoiled gradient echo (3D VIBE FLASH) sequence in the axial plane (Slices per slab = 60, Slice thickness = 2 mm, FOV = 270 mm, Matrix = 125x256). Each of these volumes, from C5 level to above hard palate, took about 18 seconds to be acquired.

Nasal and oral tracts of the Informant were also acquired while at rest (no phonation). High resolution images were obtained with a T1 weighted 3D MPRAGE sequence in the sagittal plane. Dental arches were also acquired with the informant lying in prone position and with the mouth filled with water using a 3D TSE_vfl T2 weighted sequence.

The real-time corpus was acquired using a balanced SSFP (TrueFISP) sequence (Slice thickness = 5 mm, FOV = 240 mm, Matrix = 96x128). Each word was repeated at a normal rate during 20 seconds (around 15 repetitions of each word). The resolution obtained is of 5 frames/second.

The informants were a 25 year old male (PAA) with vocal and singing training and a 45 year old female (MTC) with no vocal or singing training, both having EP as their native language. The real-time corpus and some 2D static images (for comparison with the real-time frames) were acquired with MTC. The 2D and 3D corpus were acquired with informant PAA. The reason is that, due to dental material of PAA, real-time images presented severe magnetic susceptibility artifacts which made the extraction of tongue contours impracticable. Nevertheless, we decided to maintain this informant for the static acquisitions due to its vocal and singing training, helpful to accurately sustain the phonation.

### 3. Image Processing

The 2D contours were elaborated with the Seeded Region Growing method [12]. This method was implemented and evaluated in Matlab. The results of a quantitative evaluation have shown that Region Growing is robust to changes in the placement of the seed.

![Figure 1: Adaptive grid generated for the phone [a] and respective segmentation](image-url)
computed in a similar form to [9]. In this method, we identify
the first slice (from the glottis to the mouth) where both the oral
and nasal cavities can be seen. We then choose that slice and
the next four and measure the oral and nasal area, this process
leads to five nasal and five oral areas.

4. Results

As the first examples of analyses made possible by our MRI
database, in this section we show some results regarding nasals.
We will use both 2D and 3D information.

4.1. 2D Contours

Some doubts still remain regarding nasal vowels tract configuration
in EP, MRI, by covering all tract can provide new information
in this subject. From the sagittal images we extracted
the vocal tract contours (Fig. 2), allowing their comparison. In
Fig. 3 we present some comparisons between oral and nasal
vowels. We compared each nasal vowel with the oral coun-
terparts. For nasal mid and low vowels two oral configurations
are considered.

The vowels [i] and [ɪ] present a similar configuration being
the nasal vowel produced with a higher and back position of
the tongue body and root when compared with the oral counter-
part (Fig. 3a). The velum in the nasal is lowered, allowing the
coupling with the nasal tract, as expected in the production of
a nasal vowel. As far as [u] and [u] (not shown) is concerned,
we can report that [ɪ] is farther back and slightly lower than the
oral counterpart [u], although very similar.

We can observe in Fig. 3b that the contours of the vowels
[ɛ] and [ɛ] are very close, and slightly different from [r], which
is farther back and lower.

Upon analysis of Fig. 3c, we detected some differences.
The nasal vowel [ɪ] is produced with the lowest tongue body
part in a higher position than [v] and [a]. The tongue back is
lower in [ɪ] than in [v] or [a], probably due to the lowering
velum movement.

When observing Fig. 3d, we detect that, with respect to
tongue height, the nasal vowel [o] is produced between [o] and
[a]. In the tip/blade zone, the configuration of [o] is closer to
[o] than to [a]. With regards to the tongue back, [o] it is, also,
between [o] and [a].

In this mid-sagittal images it is apparent that velum and
uvula touch the tongue back during the production of back vowels
[o] and [ʊ]. With the other nasal vowels this is not observed.

Mid sagittal distances in the pharyngeal cavity are different
in nasal vowels and their oral counterpart. As an example, [ɪ]
has a wider pharynx region relative to [v].

4.2. 3D

With MRI 3D information we could, for the first time for EP,
do the oral versus nasal vowels tract configuration comparison
using the actual configuration, represented by an area function.
Also, we were able to study actual velum port opening area. The
faster (volumic) 3D acquisition method contributes positively
to the usefulness of 3D information, by allowing much shorter
productions and slices perpendicular to the estimated tract centerline.

Area functions allowing similar comparisons to the previ-
ous section, are presented in Fig. 4.

Having the 3D information, represented in the form of area
functions, we can infer on the usefulness of the sagittal informa-
tion. In fact, the area functions generally confirm the tendencies
observed on 2D contours. Nevertheless, in 3D small or indistin-
guishable differences are clearer - for example in the pharynx
region.

Another interesting parameter for the study of the nasals is
the VPOQ. From the analysis of Fig. 5 we can observe that:
for this informant, the average VPOQ is always bigger in
the nasal vowels than in the corresponding oral ones; [ɪ]
preseñs the highest VPOQ, followed by [ʊ] and [ʊ]; the
biggest oral/nasal VPOQ difference was observed for [ʊ]/[ɪ]
show that, at least with this informant of EP, differences in the position of other articulators [9]. The 2D results showed that, at least with this informant of EP, [ɨ] is markedly higher than [a]; [ɨ] is produced with an articulatory configuration between [o] and [i]; and [i] and [u] are produced with a similar height than the oral counterparts. These results agree in general with the ones obtained using EMMA and acoustic inference from first formant values [6]. When compared to French nasal vowels, some differences are detected, particularly at the pharyngeal cavity level. French nasal vowels seem to be produced with a more constricted pharyngeal region.

With the exception of [ɨ], a central vowel that presents the biggest VPOQ, the posterior vowels have a higher VPOQ than the anterior ones. In VPOQ calculation, the oral area is always higher than the nasal, which implies a VPOQ smaller than 1, except for [ɨ] where its value is approximately one. Although the VPOQ is smaller in orals, its value is never zero. There is always a small passage to the nasal cavity. This is in agreement with the fact that nasal port opening is not sufficient to have a nasal sound.

Comparing with Engwall et al. [9], we verify that: the average VPOQ follows, in general terms, a similar behavior: superior in nasal vowels than in the correspondent orals; the VPOQ values for French are significantly higher than those obtained for EP, particularly for the nasal vowels.

5. Discussion
As expected, differences between nasals and orals are not simply summed up in the velum lowering, but there are also differences in the position of other articulators [9]. The 2D results show that, at least with this informant of EP, [ɨ] is markedly higher than [a]; [ɨ] is produced with an articulatory configuration between [o] and [i]; and [i] and [u] are produced with a similar height than the oral counterparts. These results agree in general with the ones obtained using EMMA and acoustic inference from first formant values [6]. When compared to French nasal vowels, some differences are detected, particularly at the pharyngeal cavity level. French nasal vowels seem to be produced with a more constricted pharyngeal region.

With the exception of [ɨ], a central vowel that presents the biggest VPOQ, the posterior vowels have a higher VPOQ than the anterior ones. In VPOQ calculation, the oral area is always higher than the nasal, which implies a VPOQ smaller than 1, except for [ɨ] where its value is approximately one. Although the VPOQ is smaller in orals, its value is never zero. There is always a small passage to the nasal cavity. This is in agreement with the fact that nasal port opening is not sufficient to have a nasal sound.

Comparing with Engwall et al. [9], we verify that: the average VPOQ follows, in general terms, a similar behavior: superior in nasal vowels than in the correspondent orals; the VPOQ values for French are significantly higher than those obtained for EP, particularly for the nasal vowels.

6. Conclusions
In this paper we have presented new MRI data for the EP sounds. A new and faster volume acquisition was attempted. Unlike other studies in this field, we have used a semi-automatic segmentation method. With MRI data, we can compare oral and nasal vowels contours, leading to more detailed information than previously possible with other techniques such as EMMA. Thus, we can have information on the position of the articulators, especially on the velum opening. Velum opening can be better characterized using area information from 3D MRI data, never reported before for the EP. We used VPOQ to quantify the opening areas, and it points to differences between EP and French nasal vowels. This should be studied in more detail in follow-up studies.

Future Work: At the moment, we are working on the segmentation and analysis of the remaining of the database and treatment of real-time sequences. The logic progression of this work, as we already have the ground information, is to expand the database with more repetitions and more informants. Regarding Image Processing, application of noise removal and 3D segmentation should be attempted.

7. Acknowledgements
This work is part of project HERON (POSI/PLP/57680/2004), funded by FCT (Portuguese Research Agency). The authors wish to thank the Radiology Department at the Coimbra University Hospital (HUC), particularly its Director Professor Filippe Caseiro Alves and technical staff. We also thank our two informants for their help and tolerance during acquisition session.

8. References