Velar Movement in European Portuguese Nasal Vowels

Paula Martins\textsuperscript{1}, Catarina Oliveira\textsuperscript{1}, Samuel Silva\textsuperscript{2}, and António Teixeira\textsuperscript{2}

\textsuperscript{1} ESSUA/IEETA, University of Aveiro, 3810–193 Aveiro, Portugal
\textsuperscript{2} DETI/IEETA, University of Aveiro, 3810–193 Aveiro, Portugal
pmartins@ua.pt, coliveira@ua.pt, sss@ua.pt, ajst@ua.pt

Abstract. Given the dynamic pattern of the European Portuguese (EP) nasal vowels (i.e. gradual change from an oral configuration towards a nasal one), the acquisition of information on the velum movement over time is of extreme relevance. This study aims to analyse velar dynamics during the production of EP nasal vowels by using real-time resonance imaging (RT-MRI). Audio synchronized coronal oblique slices (at velum port) were acquired from three female speakers, at a frame rate of 14 frames/s. Semi-automatic segmentation techniques were applied, in order to obtain the variation of nasal/oral areas over time, maximum nasal areas and duration of opening-closing velar gestures. The results show similar nasal areas for the different EP vowels, suggesting subtle differences in the magnitude of velum lowering. A gradual increase of nasalization over the vowel was also observed. Duration patterns reveal that the opening movement is longer than the closing gesture, which confirms previous accounts for EP.

1 Introduction

Nasal vowels are produced by a lowering of the velum, so that the nasal cavities are coupled with the oral tract. The movement of the velum during the production of nasal vowels has been studied, in great depth, over the last decades using a varied set of direct (e.g. fiberoscopy [1], cineradiography [2], electromagnetic articulography [3], magnetic resonance imaging [4]) and indirect techniques (e.g. measurements of oral and nasal air flows [5], acoustic analysis [6]). Most of this research has focused on French nasal vowels, but differences between Portuguese nasalization and nasalization in other languages warrant specific studies of Portuguese nasal sounds.

The articulation of European Portuguese (EP) nasal vowels has received a considerable amount of attention by our team, by exploring electromagnetic articulography (EMA) data [7–10] and, more recently, magnetic resonance imaging (MRI) techniques [11, 12].

EMA studies on EP nasal vowels examined essentially duration, velocity (stiffness), magnitude and timing of velum gestures and suggest the following conclusions: i) the opening and closing movement of the velum takes about 220ms (fast speech rate) - 340ms (normal speech rate) and the opening movement
is longer than the closing gesture \([13, 10]\); ii) stiffness differences between the opening and the closing gestures of the velum entail differences on the duration of articulatory movements \([10]\); iii) the velum gesture is somewhat delayed with respect to the tongue body gesture, inducing an oral onset, and depending on the degree of overlap between the closing velum gesture and the following oral gesture (as in “\textit{conto}”), a transitional/intrusive consonantal segment can appear \([8, 10]\); iv) speech rate reduces the duration of velum gestures and increase the stiffness and the intergestural overlap \([10]\); and v) as previous studies on velum height for nasal segments \([7, 9]\) provided some inconclusive or ambiguous results, further data should be acquired to find whether the velum position is identical between the five EP nasal vowels.

Static 2D MRI has been used to obtain information about articulatory differences in the tongue and velum position between oral and nasal vowels, through superimposition of midsagittal contours \([11, 12]\). 3D MRI allowed us to measure nasal and oral cross-sectional areas and to calculate the velum port opening quotient (VPOQ) \([11, 12]\).

While EMA provides high temporal resolution measurements of articulator movement, although limited to a few points, MRI gives accurate information on the position of the articulators, such as tongue and velum, but with some well-known drawbacks, especially long acquisition times, requiring the subject to artificially sustain the articulation. Thus, both EMA and static MRI have to be complemented with other methods, such as real-time MRI (RT-MRI), in order to fully characterize articulatory movements. Given the dynamic pattern of Portuguese nasal vowels (i.e. gradual change from an oral configuration towards a nasal one), first mentioned by Lacerda and Head \([14]\) and underlined also by perceptual \([15]\) and articulatory studies \([8, 10]\), the acquisition of dynamic information is particularly relevant.

This study complements previous articulatory analysis by using RT-MRI to analyse temporal course of velar movement during the articulation of EP nasal vowels by three EP speakers. The present paper reports, for the first time, variation of nasal and oral areas through time, as well as duration of velar opening and closing movements measured from the nasal areas. Furthermore, potential differences in the magnitude of velum opening between the five EP nasal vowels will be examined, in order to establish if there is more than one position of velic lowering (cf. Demolin et al. \([16]\) for French nasal vowels).

Information concerning areas and durations is crucial to realistically represent the movement of articulatory gestures in our articulatory based text-to-speech system \([17]\).

This paper is organized as follows: section 2 details methods of MR image acquisition and describes the tools used for image processing and analysis; section 3 provides some results; finally, in section 4, we briefly discuss our results and summarize directions for future work.
2 Methods

2.1 Corpus and Speakers

Three female speakers (SV, CM, CO) of EP, aged between 21 and 33, participated in the RT-MRI experiment. None of the speakers reported having hearing or speech disorders. An MRI screening form and informed consent was obtained before the study to comply with security and ethics rules.

The corpus considered in this study represents a subset of a large database (refer to [18] for more details).

The speakers produced nonsense words containing EP nasal ([ɪ̊̊̊̊], [œ̊̊̊̊], [ɪ̊̊̊̊], [o̊̊̊̊], [ʊ̊̊̊̊]) vowels, uttered in three prosodic conditions: word-initial, word-internal and word-final (e.g. “ampa, pampa, pan”, [œ̊̊̊̊],[pœ̊̊̊̊],[pʊ̊̊̊̊]). The nasal vowels were flanked by a voiceless bilabial stop [p].

Audio was recorded simultaneously with the RT-images, inside the MR scanner, at a sampling rate of 16000 Hz, using a fiberoptic microphone.

2.2 Image Acquisition

The MRI experiment was carried out in a Magnetic Resonance Imaging Unit at Coimbra (IBILI – Institute of Biomedical Research in Light and Image). The images were acquired on an unmodified 3.0 T MR scanner (Magnetom Tim Trio, Siemens, Erlanger, Germany) equipped with high performance gradients (Gmax = 45mT/m, rise time = 0.2s, slew Rate = 200 T/m/s, and FOV = 50 cm). A 12-channel head and 4-channel neck phased-array coils were used for data acquisition. Parallel imaging (GRAPPA 2) together with magnetic field gradients operating at FAST mode were used to speed up the acquisition. The subject lay supine in the MR scanner, while producing the stimuli and wore headphones to protect the ears from the noise. The imaging protocol was an evolution of a former study conducted by Martins et al. [19].

After localization images, a T1 W 2D-coronal oblique MRI slice was taken at the velum, using an Ultra-Fast RF-spoiled Gradient Echo (GE) pulse sequence (Single-Shot TurboFLASH), with a slice thickness of 8 mm and the following parameters: TR/TE/FA = 72ms/1.02ms/5°, Bandwidth = 1395 Hz/pixel, FOV(mm²)= 210 × 210, reconstruction matrix of (128 × 128) elements with 50 % phase resolution, in-plane resolution (mm²) = 3.3 × 1.6, yielding a frame rate of 14 images/second (i.e. temporal resolution of 72ms). The acquisition of each sequence took about 5 seconds, resulting in 75 coronal oblique images. A previously obtained sagittal slice was used to better determine the orientation of the oblique slice.

Figure 1 provides details concerning the acquired image sequences depicting the oblique acquisition plane and the location of the oral and nasal cavities.

2.3 Data Processing

Data processing consists in three main steps: image segmentation, to identify the oral and nasal cavities and compute their areas for all image frames; audio
Fig. 1. From left to right: mid-sagittal plane depicting orientation of the oblique plane used during acquisition, sample oblique plane showing the oral and nasal cavities and image sequence details.

Fig. 2. Data processing pipeline for image and audio data.

Annotation, to be able to identify the set of frames corresponding to the nasal vowels; and, finally, analysis of the extracted curve segments. Figure 2 depicts the data processing pipeline.

Since the subject keeps the same head position, throughout the acquisition, the oral and nasal cavity are located in the same region in every image. This was explored in order to speed segmentation of both cavities along the image sequence (75 frames).

A region growing method was used, defining a neighbourhood for each seed including both spatial coordinates and time. This allowed that, with a small number of seeds, each of the cavities could be segmented along the whole image sequence.

Segmentation starts by the oral cavity, where the user has to position a seed. The intensity interval for the region growing (automatically established from the seed neighbourhood) can be tuned for each image sequence but a value around 6% of variation has provided good results overall. The number of seeds needed to fully segment the oral cavity depends on the number of times the cavity closes during the sequence. For the processed sequences only one seed was required to fully segment the oral cavity.

The nasal cavity was segmented using a similar method to the one used for the oral cavity. Since the nasal cavity closes several times along each image sequence, five seeds (in different image frames) were typically needed to perform the segmentation. Further details concerning the segmentation method can be found in Silva et al. [20].
Figure 3 shows sample frames depicting the oral and nasal cavities segmentations. For this particular case the user defined a total of six seed points: one for the oral cavity (in frame 16) and five for the nasal cavity (in frames 0, 25, 28, 37 and 56). Segmentation time was around two minutes for each image sequence.

Audio segmentation was performed manually, using the software tool Praat [21], in order to locate the target sounds. Since the audio is synchronized with the image sequence it is possible to identify the set of image frames corresponding to the production of each vowel in the different prosodic contexts.

Velar opening, plateau and closing durations were computed from the nasal area curves. The first derivative for the nasal area curve segments was computed and passings through 10% of the derivative maximum used as landmarks for the transition between gestures (as illustrated in figure 4).
3 Results

3.1 Nasal/oral area variation over time

Figure 5 presents, as a representative example of what can be observed for all speakers, nasal and oral areas, over time, for all EP nasal vowels produced by speaker CO in three different word positions flanked by the plosive [p].

In the beginning of the nasal vowel, nasal areas (N) are very small (lower than 25 mm²) denoting a small velum port opening (VPO). After that, the curves show a gradual increase of the areas until reaching the maximum nasal area (lowest velar position) and, finally, the areas decrease again for the production of [p]. The magnitude of the nasal areas is quite similar for the different vowels (approx. 100 mm² for this speaker). This pattern is observed both in word-initial and word-internal position. In word-final position, after reaching the maximum nasal area, the area does not decrease since the velum remains open.

The oral areas (O) follow an inverse trend, decreasing while the nasal area increases. However, the presented curves show differences in oral areas between the five EP vowels. Oral areas are generally higher for [ı] and [i] than for [o] and [u] in consequence of different tongue configurations among the vowels.

3.2 Velar Port Opening Areas

Figure 6 shows the mean for the maximum nasal areas for each speaker and EP nasal vowel (left) in different word-positions (right).
Fig. 6. Left, average of the maximum nasal areas for each of the EP nasal vowels produced by the 3 speakers in all positions considered in this study; right, average of the maximum nasal areas for each of the three word positions.

Fig. 7. Mean duration of velar opening, plateau and closing for speakers CO and CM.

Comparing the results obtained, differences in the nasal areas between the three subjects can be observed. Subject CO presents the highest nasal areas (between 102.2 mm$^2$ for [e] and 116.8 mm$^2$ for [i]), followed by subject SV (between 84.1 mm$^2$ for [e] and 98.5 mm$^2$ for both [e] and [u]) and, finally, subject CM (between 79.3 mm$^2$ for [e] and 86.8 mm$^2$ for [u]). This reflects anatomical differences between subjects, though not very accentuated, as they are all female with similar biotype.

Figure 6, on the left, suggests only slight differences in nasal areas among vowels, for all speakers. The observed differences in VPO do not seem to be related with differences in tongue position of the vowels.

As illustrated in Figure 6, on the right, VPO areas of the nasal vowels produced in different word positions are rather similar to each other, at least for CO and CM. Subject SV reveals a difference of 23 mm$^2$, between initial and medial word position.

3.3 Duration of velum movements

Figure 7 shows the mean durations (in ms) of velum opening ($d_{\text{Open}}$), plateau ($d_{\text{Plateau}}$) and closing ($d_{\text{Close}}$) gestures of EP nasal vowels computed for two of the speakers, CO and CM.

Mean total duration of velar opening - closure gesture is 437 ms for speaker CO and 333 ms for speaker CM. As shown in figure 7, mean $d_{\text{Open}}$, $d_{\text{Plateau}}$ and $d_{\text{Close}}$ are respectively 203.63 ms/43.88 ms/164.25 ms for speaker CO and 162.00
ms/28.13 ms/131.63 ms for speaker CM. For both speakers, opening gesture is longer than closing movement.

The $d_{\text{Open}}$ variation pattern, for each vowel, is similar for the two analysed speakers: short $d_{\text{Open}}$ for [ʊ], followed by [ɪ], [ı] and [o], with similar duration values, and finally [e], presenting the longest duration. For all vowels, $d_{\text{Open}}$ is longer than $d_{\text{Close}}$, except for [ʊ].

### Table 1. Mean values of observed maximum oral-nasal areas proportion, $P_{\text{NO}}$, by speaker and vowel.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>[ʊ]</th>
<th>[e]</th>
<th>[ı]</th>
<th>[o]</th>
<th>[u]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO</td>
<td>0.47</td>
<td>0.66</td>
<td>0.39</td>
<td>0.60</td>
<td>0.97</td>
</tr>
<tr>
<td>SV</td>
<td>0.30</td>
<td>0.30</td>
<td>0.35</td>
<td>0.76</td>
<td>0.99</td>
</tr>
<tr>
<td>CM</td>
<td>0.39</td>
<td>0.47</td>
<td>0.34</td>
<td>0.96</td>
<td>0.99</td>
</tr>
</tbody>
</table>

### 3.4 Nasal-oral areas proportion

The nasal-oral proportion, $P_{\text{NO}} = N/(N + O)$, was computed, as well as its variation over time. As expected, the variation of the $P_{\text{NO}}$ over time follows a similar pattern to the one observed for the nasal areas. At the beginning of the nasal vowels, $P_{\text{NO}}$ is nearly zero, increasing over vowel production and decreasing again for [p] articulation.

However, the magnitude of the $P_{\text{NO}}$ differs between nasal vowels. Since the nasal areas are similar among vowels, as previously shown, this variation is mostly due to differences in oral areas.

The mean maximum values of $P_{\text{NO}}$ by speaker and vowel are shown in Table 1. For all the speakers, the $P_{\text{NO}}$ is higher for back vowels ([o] and [ʊ]) than for the more fronted ones ([ɪ], [e] and [ı]).

### 4 Conclusions

The results obtained in this study corroborate previous accounts for EP, namely that nasality in Portuguese is typically incremental over the vowel, with a movement from oral to nasal [14, 10]. This trend was consistently observed for the three speakers, in all vowels and word positions. As expected, in word-internal position, nasal vowel is partly produced as an oral vowel (i.e. nasal areas are very small), since the nasal vowel follows a plosive, that only allows slight velopharyngeal port openings [22]. However, in initial position, small nasal areas were also observed, that reflect a closing movement of the velum, when it would be predicted a continuity of the velum opening from the rest position.

Moreover, nasal area values suggest that the magnitude of the velopharyngeal port opening is similar between the different vowels. These findings are in agreement with previous EMA data, obtained by Oliveira [9], for one speaker, that indicated there were no significant differences in velum height between nasal vowels. Although velum height and VPO are different measures, they are correlated [23].
The mean durations of velum opening, plateau and closing are, in general, consistent with previous EMA data [13, 10]. The opening movement is longer (25–35 ms) than the closing movement, for both speakers and vowels, excluding the high back nasal vowel ([i]).

In sum, oblique RT-MRI slices at the velopharyngeal port provided new articulatory data on time varying nasal areas, that complement and clarify previous results obtained with EMA, allowing for a better characterization of velum movement in the production of nasal vowels.

The available image data might also allow the study of velar port configuration (shape) over time (figure 8), another important feature for proper modelling of EP nasal vowels. Furthermore, temporal organization of gestures will be addressed, through the analysis of real-time image sequences acquired both at the velum and lips.

Acknowledgements

This research was partially funded by FEDER through the Operational Program Competitiveness factors (COMPETE) and by National Funds through Foundation for Science and Technology (FCT) in the context of the Project HERON II (PTDC/EEA-PLP/098298/2008). The first author acknowledges the PhD grant from FCT (SFRH/BD/65183/2009). The authors thank Professor Miguel Castelo-Branco (IBILI Director) and the speakers involved in the study.

References