Segmentation and Analysis of the Oral and Nasal Cavities from MR Time Sequences

Samuel Silva[1], Paula Martins[2], Catarina Oliveira[2], Augusto Silva[1], and António Teixeira[1]

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

Abstract. The study of dynamic aspects of speech production in Portuguese is very important to characterize vowel nasalization. In this context, the analysis of velum movement remains a challenging task and only a few studies present articulatory descriptions of Portuguese nasal vowels.

Advances in real-time MRI (magnetic resonance imaging) allow the acquisition of vocal tract images with reasonable spatial and temporal resolution to enable observation and quantification of articulatory movements. The resulting data consists of large image sequences and the structures of interest (e.g., oral cavity) have to be identified (segmented) throughout to enable analysis which can be a time consuming task.

This article presents a segmentation tool for real-time MR image sequences of the oral and nasal cavities. The proposed tool has been implemented using MevisLab and provides features for the analysis of the resulting data.

1 Introduction

Vowel nasalization is one of the most important features of Portuguese language. It has five nasal vowels [ɐ̃, ẽ, ĩ, õ, ũ] (e.g., "c**an**tar" ("to sing")) and several nasal diphthongs (e.g. "c**ão**" ("dog")). The production of a nasal vowel involves flow of air through both the mouth and the nose. The access for the airstream to the nasal passages is controlled by the position of the velum: to put it simply, when the velum is lowered and the velopharyngeal port is open, the airstream escapes through the nose and the sound is nasal; oral sounds are produced with the velum raised, so that the access to the nasal cavities is blocked.

Several studies on Portuguese nasal vowels revealed that the initial segment of the nasal vowel is basically oral or slightly nasalized and the nasality increases progressively over the vowel's duration [1, 2]. Many researchers also took notice of the emergence of a final consonantal segment ("nasal tail") after the nasal vowel, essentially in preconsonantal contexts (e.g. "c**am**po" ("field")) [3–6] and less often in final position (e.g. "f**im**" ("end")) [7]. Teixeira et al. [8] highlighted the importance of velar dynamics in the perception of synthesized nasal vowels.

In this sense, the study of dynamic aspects of speech production in Portuguese is of the utmost importance. However, the analysis of velum movement is a very challenging task and the few recent studies that have provided articulatory descriptions of Portuguese nasal vowels had several limitations: some of them concern Brazilian Portuguese varieties [3]; in general, data come from a small number of speakers and contexts [9, 6]; articulatory techniques used, such as EMA [9, 6] or fiberscope [3], are quite invasive; EMA only provides dynamic information of a few points across time [9, 6]; and static MRI allows complete visualization of all the vocal tract organs but does not provide dynamic information. [10, 11].

Recent advances in MRI technology have allowed real-time images of the vocal tract with reasonable spatial and temporal resolution in a non-invasive manner. These advantages make real-time MRI suitable for the observation and quantification of articulatory movements, as well as visualization of the vocal tract shape. Considering that such dynamic studies result in a large amount of data (images), systematic (and not very time consuming) methods should be devised to process it in order to extract relevant articulatory features such as areas of oral and nasal cavities, volumes and distances. Most of the times, the first approach should provide a quick insight, allowing exploration of the data set, which will then guide new lines of work (and, eventually, more elaborate data analysis). Furthermore, the developed tools should be adequate for users with different backgrounds (e.g., linguists and radiographers).

We present a segmentation method for the oral and nasal cavities from real-time MR image sequences providing a fast processing of the whole image sequences in 2-3 minutes. The method was developed in MeVisLab and provides a graphical user interface which works as a frontend to the implemented processing network. Some simple analysis features are also provided.

This article is organized as follows: section 2 describes data collection; in section 3 the segmentation method is presented followed by a discussion of its application to several image sequences. Finally, section 5 presents some conclusions and ideas for future work.

2 Data Collection

2.1 Corpus and Speakers

Data was acquired for a female speaker, aged 33, phonetically trained, with no history of hearing or speech disorders. Each image sequence was acquired while the subject was saying nonsense words containing the five European Portuguese (EP) nasal vowels in three prosodic conditions: word-initial, word-internal and word-final (e.g. “ampa, pampa, pan”). The nasal vowels were flanked by the bilabial stop ([p]) or the labiodental fricative ([f]). Each stimulus set was prompted orally by one of the experimenters over the intercom system.

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

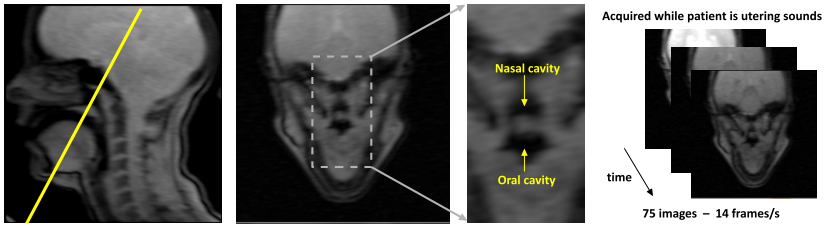

Fig. 1. From left to right: mid-sagittal plane depicting orientation of the oblique plane used during acquisition, sample oblique plane showing the oral and nasal cavities and image sequence details.

2.2 Image Acquisition

The MRI experiment was carried out in a Magnetic Resonance Imaging Unit at Coimbra (Institute of Biomedical Research in Light and Image). The images were acquired on an unmodified 3.0 T MR scanner (Magneton 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 custom 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. An MRI screening form and informed consent was obtained before the study to comply with security and ethics rules. The subject lay supine in the MR scanner, while producing the stimuli and wore headphones to protect the ears from the noise.

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 = $72ms/1.02ms/5^o$, Bandwidth = 1395 $Hz/pixel$, FOV(mm^2)= 210 x 210, reconstruction matrix of (128 x 128) elements with 50 % phase resolution, in-plane resolution (mm^2) = 3,3 x 1,6, yielding a frame rate of 14 images/second. 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.

3 Image Segmentation

Selected frames from one of the acquired MR image time sequences were previously segmented using a Live-wire [12] based approach with a radiographer identifying the contours of the oral and nasal cavities for each frame individually [13] which is a time consuming method if applied to the whole sequence.

Our main goal was to provide a quick segmentation of the image sequences in order to allow easier exploration of the image data.

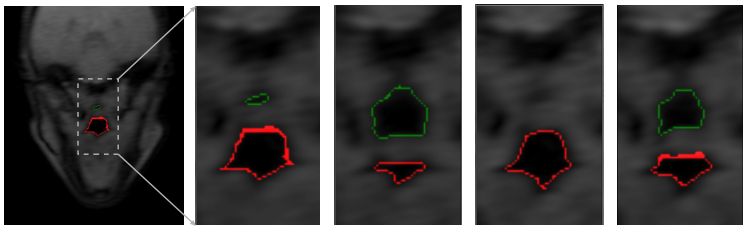

Fig. 2. Different frames depicting the segmented oral and nasal cavities.

A description of the devised method and its implementation are provided in what follows.

3.1 Method

One important aspect concerning the obtained images is that, for each time sequence, the subject keeps the same head position throughout the acquisition, which results in spatial coherence between frames. Therefore, the oral and nasal cavity are located in the same region in every image. Our main idea was to explore this characteristic in order to speed segmentation of both oral and nasal cavities for the 75 images in each sequence.

A region growing method was used defining a 4D neighbourhood for the seeds, considering spatial coordinates and time. This allowed that with a small number of seeds (depending on the number of times each cavity completely closes during the sequence, which stops the region growing) each of the cavities could be segmented.

Segmentation starts by the oral cavity. The user has to position a seed inside the oral cavity, desirably in a frame where the cavity is well defined (or has a considerable size), since this allows to position the seed in a region with a clearly low intensity, a proper initialization for the region growing. The intensity interval for the region growing can be tuned for each image sequence but a value around 6% has provided good results overall.

The nasal cavity is segmented using a similar method to the one used for the oral cavity. Since the nasal cavity is generally smaller than the oral cavity (and not as clearly hypo-intense), defining a suitable seed might not be as easy. This might result, for example, that the region growing overspills to the oral cavity. To tackle this problem the allowed intensity interval for the region growing might be narrowed but this often results in a poorer segmentation overall when, most of the times, the problem arises only due to one or two time frames. Therefore, before nasal cavity segmentation the previously obtained oral cavity segmentation is masked with a high intensity value preventing overspill.

Figure 2 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).

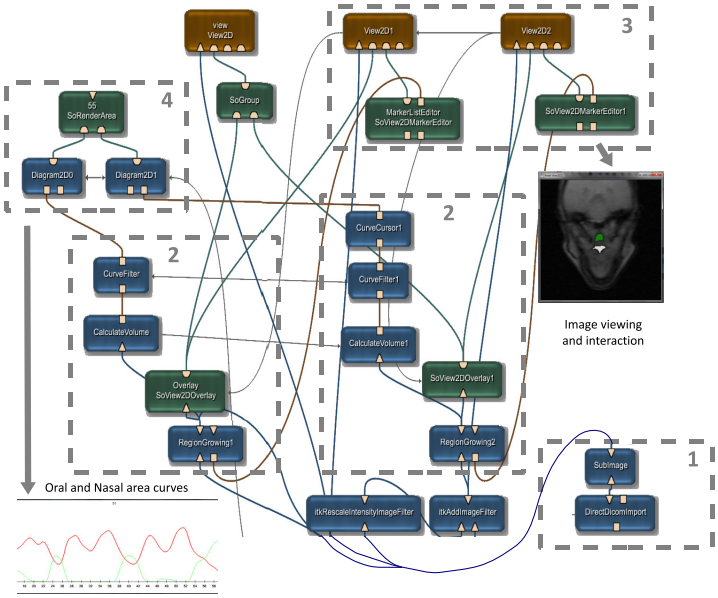

Fig. 3. Module network built in MevisLab: 1) image loading; 2) region growing method; 3) image visualization and interaction (seed definintion) for each cavity; 4) oral and nasal area presentation over time.

3.2 Implementation

The segmentation method was implemented using MeVisLab (MeVis Medical Solutions, `http://mevislab.de`). This software tool allows building image processing pipelines by connecting processing/visualization modules in a network.

Figure 3 shows the module network for the described segmentation method encompassing modules for image loading, region growing management and image visualization.

To access image viewing windows or changing parameters (e.g., intensity interval for the region growing) the user just needs to double click the corresponding module. Nevertheless, in order to provide a more user friendly experience, since this tool will be used by users with different backgrounds (radiologists, linguists, etc.), a simple graphical user interface has been designed gathering only the relevant features for the segmentation tasks to perform. Figure 4 shows the different aspects of the graphical user interface supporting image loading, seed definition, output visualization and detailed analysis.

4 Results and Discussion

The segmentation method described above was applied to image sequences for the different nasal vowels ([ɐ̃, ẽ, ĩ, õ, ũ]) in [p] context. The segmentation was

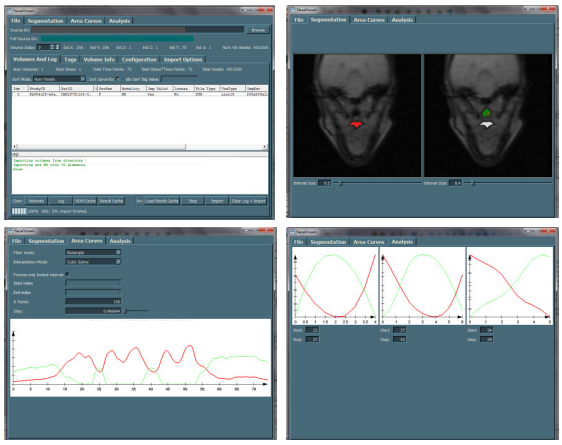

Fig. 4. Simple graphical user interface, built in MeVisLab, to support oral and nasal cavities segmentation. From top to bottom and left to right: panels for image loading; segmentation of oral and nasal cavities; area curves visualization; and analysis of subsets of frames corresponding to a specific sound.

performed in less than three minutes, for each sequence, including seed definition and segmentation inspection. In general, five to seven seeds had to be defined (usually one or two for the oral cavity and the remaining for the nasal cavity).

The segmentations allow the computation of oral and nasal cavity areas over time. Figure 5 shows the area curves for the oral and nasal cavities for three EP nasal vowels in different word positions.

Nasal area progressively increases over the vowel and decreases for the production of the consonant ([p]), while for the oral area is precisely the opposite (see figure 6). Regardless the vowel and its word position, nasal areas are consistently smaller than the oral ones. If confirmed with more data and speakers, this trend distinguishes Portuguese from other languages with nasal vowels in their inventory of sounds, such as French [14]. Also, nasal areas are identical between the five different EP vowels, which is in agreement with a previous study based on EMA data [15]. On the contrary, oral areas appear to vary with vowel height. An oral/ nasal ratio can be computed from these areas, with important implications for cross-linguistic investigations of vowel nasality.

Furthermore, as described earlier, sound was also recorded during image acquisition. By analysing the sound files it is possible to match sets of frames to specific sounds. For example, figure 6 shows the area curves for the oral and nasal cavities while the subject was uttering the nasal vowel [ɐ̃] in different word positions ("ampa, pampa, pan") in [p] context. Notice how the variation of oral and nasal cavity areas are different for the different word positions specially for "pan".

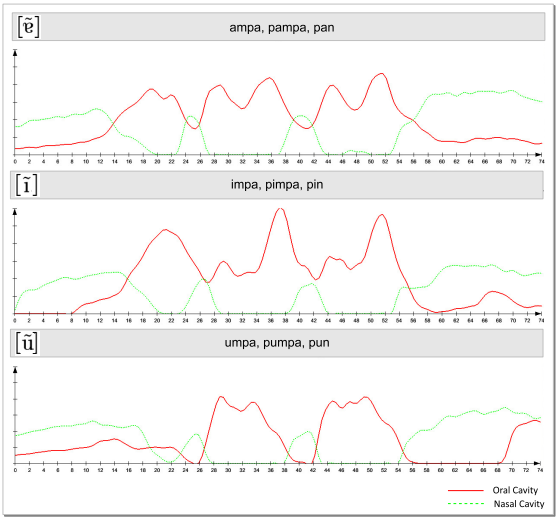

Fig. 5. Curves depicting the area for the oral and nasal cavities for nasal vowels [ɐ̃, ĩ, ũ] in different word position in [p] context.

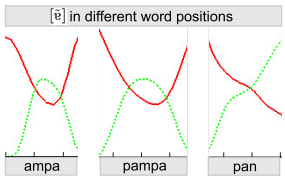

Fig. 6. Areas for the oral and nasal cavities corresponding to the nasal vowel [ɐ̃] as in "ampa", "pampa" and "pan".

5 Conclusions and Future Work

This article presents a segmentation method for MR time sequences of coronal-oblique slices encompassing the oral and nasal cavities. The devised method provides a simple and fast computation of oral and nasal cavities areas and presents the user with a line graph depicting their values over time. The simple graphical user interface developed in MeVisLab provides a high level interface with the module network showing only relevant windows and parameters. Considering that the analysis of this time sequences might need to concentrate in specific frames to isolate specific sounds (based on the analysis of the audio recorded during image acquisition) a feature is also available to show only a subset of the frames for comparison.

Expanding the corpus with data from additional speakers is fundamental to account for inter-speaker differences and will provide stronger evidence for the tendencies observed in the discussion. In this scenario the presented tool is an important asset, allowing a systematic and fast approach to the resulting image

dataset. Data from additional speakers will also enable further validation of the proposed segmentation method encompassing different aspects such as intra and inter-observer variability.

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 second 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

1. Sampson, R.: Nasal Vowel Evolution in Romance. Oxford University Press (1999)
2. Lacerda, A., Head, B.: Análise de sons nasais e sons nasalizados do português. Revista do Laboratório de Fonética Experimental de Coimbra **6** (1966) 5–70
3. Lovatto, L., Amelot, A., Crevier-Buchman, L., Basset, P., Vaissière, J.: A fiberscopic analysis of nasal vowels in Brazilian Portuguese. In: ICPhS' 2007, Saarbrücken (2007) 549–552
4. Sousa, E.M.d.: Para a caracterização fonético-acústica da nasalidade no Português do Brasil. Msc dissert. in linguistics, Univ. Est. de Campinas, Brasil (1994)
5. Ashby, S., Cummins, F., Barbosa, S., Campani, N.: Nasal heads or nasal tails? a pan lusophone survey of preconsonantal murmur. In: ICPhS 2011, Hong Kong (2011)
6. Oliveira, C., Martins, P., Teixeira, A.: Speech rate effects on European Portuguese nasal vowels. In: InterSpeech, Brighton (2009)
7. Shosted, R.: Excrescent nasal codas in brazilian portuguese: an electropalatographic study. In: ICPhS 2011, Hong Kong (2011)
8. Teixeira, A., Vaz, F., Príncipe, J.: Influence of dynamics in the perceived naturalness of Portuguese nasal vowels. In: ICPhS, San Francisco (1999)
9. Oliveira, C., Teixeira, A.: On gestures timing in European Portuguese nasals. In: ICPhS, Saarbrücken (2007)
10. Martins, P., Carbone, I., Silva, A., Teixeira, A.: An MRI study of European Portuguese nasals, Antwerp (2007)
11. Martins, P., Carbone, I., Pinto, A., Silva, A., Teixeira, A.: European Portuguese MRI based speech production studies. Speech Communication **50**(11–12) (2008) 925–952
12. Falcão, A., Udupa, J., Samarasekera, S., Sharma, S., Hirsch, B., Lotufo, R.: User-steered image segmentation paradigms: Live wire and live lane. Graphical models and image processing **60**(4) (1998) 233–260
13. Teixeira, A., Martins, P., Oliveira, C., C., F., Silva, A., Shosted, R.: Real-time MRI for Portuguese: Database, methods and applications. In: Proc. PROPOR 2012, Coimbra (2012)
14. Engwall, O., Delvaux, V., Metens, T.: Interspeaker variation in the articulation of French nasal vowels. In: Proc. 7th ISSP, Ubatuba (2006) 3–10
15. Oliveira, C.: Do Grafema ao Gesto: Contributos Linguísticos para um Sistema de Síntese de Base Articulatória. PhD thesis, University of Aveiro (2009)