

Investigating the Differences Between European Portuguese Sounds: An Approach using Polygonal Mesh Comparison

Paula Martins^{1,2}, Samuel Silva^{2,3}, Catarina Oliveira^{1,2}, Carlos Ferreira⁴, Augusto Silva^{2,3}, António Teixeira^{2,3}

¹Escola Superior de Saúde, Universidade de Aveiro

²IEETA – Instituto de Engenharia Electrónica e Telemática de Aveiro

³Dep. Electrónica, Telecomunicações e Informática, Universidade de Aveiro

⁴ANIFC – Associação Nacional de Imagiologia Funcional e Cerebral

ABSTRACT

With the main purpose of studying the articulatory configuration of the European Portuguese (EP) sounds and to evaluate coarticulatory effects, a large 3D MRI database from several speakers and contexts have been acquired. Segmentation, visualization and interpretation of all the available data, essential to a detailed characterization of these sounds, are complex and time-consuming tasks. To perform the segmentation, a semi-automatic method was used. Meshes of the tongue were created from the segmented volumes and a polygonal mesh comparison tool (PolyMeCo) was used to compare: different sounds, the same sound produced in different contexts and syllabic positions, and productions from different speakers, providing a qualitative measure of these differences. To the best of our knowledge this is the first study in the literature using such an approach to analyze and compare tongue shape. Furthermore, these first results are very promising allowing a proper analysis and depiction of the main differences between EP sounds and further insight into their production.

1. Introduction

Functional and morphological modeling of the vocal tract structures is of special relevance in the fields of speech production, speech synthesis and speech disorders. Biomechanical models of the oral, laryngeal and pharyngeal structures have been developed, in the last years, with a relevant contribution, not only for a better understanding of speech production mechanisms (e.g. [PPZP03]; [GPP06]), but also for the diagnoses and treatment of speech and sleep disorders (e.g. [FSH*09]). In the speech synthesis field, articulatory synthesis [BK06] is one of the most promising techniques [SD02], but this type of anthropomorphic synthesizers demands large amounts of detailed anatomic-physiological information, if possible in 3D [TMS*05]. From a phonetic/linguistic point of view, articulatory information and 3D models are essential to improve knowledge concerning the articulation of the different sounds and can also be used as an auxiliary tool when teaching e.g. foreign languages.

1.1. Instrumental techniques

Obtaining accurate and reliable articulatory information from the different vocal tract articulators involved in speech production [NAH97] is quite challenging. Several and increasingly accurate instrumental techniques of measuring vocal tract configurations have emerged in recent years.

Amongst them, different imaging modalities (X-Ray, computed tomography (CT), magnetic resonance imaging (MRI) and ultrasound) have been used with the main goals of obtaining articulatory data, during speech production, and also for modeling purposes. However, the use of Cineradiography and CT raised ethical concerns due to the use of ionizing radiation, precluding its utilization in speech production studies, in healthy subjects. Ultrasound was used, e.g., to obtain tongue dorsum surface shapes [SL96], to evaluate lingual coarticulation [ZHH08] and has recently received particular attention in Silent-Speech interfaces [HAC*07]. Although providing an excellent temporal resolution, the information obtained is restricted to the dorsum of the tongue.

MRI has some potential advantages in speech production field: it provides good contrast between soft tissues (e.g., tongue, velum), allows 3D modeling (particularly relevant in the production of lateral sounds) and has the capacity of gathering information from the vocal tract in all of its extension [NAH97]. This last advantage is of special interest in the study of the pharyngeal cavity, since it is not accessible through other articulatory techniques. Moreover, it is a non-invasive and safe imaging technique that does not involve the use of ionizing radiation. MRI limitations in the field are well known, including: difficulties in observing calcified structures (e.g., the teeth), the acquisition technique, in which the speaker must be lying down during speech

production (influence in tongue posture) and the inherent noise produced during the acquisitions (expensive technology is required to record acoustic signal) [NAH97]; [SD02]; [EB99]. The relatively low temporal resolution achieved, even with the fastest acquisition techniques, is also a constraining factor if dynamic information is needed. However, some of these limitations have been partially overcome, with recent technological advances in the field, increasing the applicability of this imaging technique in this area. For other languages (e.g., French, English, Swedish, German, etc.) several MRI studies [BBRS98, NAH97, EB99] have been performed to obtain articulatory data or attain modelation of vocal tract organs and sounds.

1.2. European Portuguese

For European Portuguese (EP), there is a lack of articulatory data, particularly, but not only, for specific classes of sounds (e.g., laterals, rhotics) [Oli09]. In the last few years a great investment in this field has been performed [MCP*08, TOM*09] with the main goal of developing an articulatory synthesizer (SAPWindows) [TMS*05]. At this moment, oral vowels, nasal vowels and fricatives can be synthesized, with good results. Synthesis of other sound classes, such as laterals, will require 3D articulatory information (e.g., vocal tract areas, area and extension of the lateral channels) and dynamic data (e.g. nasal vowels, rhotics and diphthongs).

In this study, we use MRI to study articulatory characteristics of some EP sounds with the main focus on lateral consonants. EP has two lateral consonants: the /L/ (e.g. *palha*, *straw*) and /l/ (e.g. *sal*, *salt*). The /L/ is produced by one movement of the anterior tongue dorsum against the alveolo-palatal area. The production of /l/ involves linguo-alveolar contact and the establishment of one or two lateral passages along the sides of the tongue. However, some interspeaker variability in the /l/ production has been reported and different patterns were found amongst different languages and dialects. In some languages, there are differences in the /l/ configuration depending on its position in the syllable: the /l/ produced at the end of a syllable has a more retracted tongue dorsum or elevation towards the velum, than the /l/ produced at the beginning of the syllable. The latter is called *light* or non velarized /l/ and, the former a *dark-l* [RE05].

For EP it is not clear if these differences, on /l/ production, occur or not. Previous acoustic [And99] and MRI data [MCP*08] seem to confirm the theory of the existence of a single configuration, in all syllabic positions. Moreover, it was not clear, for EP, how much the /l/ (and other sounds) are influenced when produced in the context of different vowels. The articulatory or acoustic influence of one speech segment on another is called coarticulation and is known to occur universally, but varies from one language to another [FR99]. Recent theories of speech production consider that coarticulation plays a crucial role in speech and is essential to take it into account in speech production models and speech synthesis.

At present, our main goals are the exploration of the 3D data, allowing an articulatory description of the sounds, and the extension of the results already obtained from 2D data. While, for some classes of sounds, 2D data (usually in the midsagittal plane) provides a rather good characterization of these sounds, for lateral sounds 3D information is crucial. To

attain these goals it is important to find efficient segmentation techniques and explore different tools and metrics that might contribute for a deeper insight into the production of EP sounds.

In the next section the image acquisition technique is briefly presented followed by a description of the devised semi-automatic segmentation method. A short description of the tool used for comparison is also included. Relevant examples of the different comparisons performed are presented and discussed in section 3. Finally, section 4 presents the main conclusions along with ideas for further work.

2. Image Acquisition, Segmentation and Polygonal Mesh Comparison

2.1. Image acquisition

MRI acquisition was carried out in a Magnetic Resonance Imaging Unit at Coimbra (Institute of Biomedical Research in Light and Image (IBILI)) using a 3.0 T MR scanner (Magnetom Tim Trio, Siemens, Erlanger, Germany) equipped with high performance gradients ($G_{max}=45\text{mT/m}$, rise time=0.2s, slew rate=200 T/m/s; and FOV=50 cm). A standard 12-channel Head and Neck phased-array coils and parallel imaging (GRAPPA) were used in all data acquisition sessions. The imaging protocol used in the present study was based in a previous MRI study [MCP*08]. The subjects were positioned comfortably in a supine position using headphones. After acquiring reference images, a T1 W 5 mm thickness midsagittal MRI slice of the vocal tract was obtained using a TSE sequence [BKZ04] (TR/TE/FA=400 ms/7.8 ms/120°), FOV=240x240 mm; matrix (256x256) resulting in a pixel size of (0.938, 0.938). The acquisition time was 6 seconds. After that, a volume covering the entire vocal tract was obtained in the sagittal plane with a T1W 3D Spoiled GE sequence (VIBE), resulting in an acquisition time of 19 seconds; matrix (224x256); voxel size (1.055, 1.055, 2). The speakers sustained the sound during acquisition; the sequence was launched when the sound was produced (e.g., *sallllllll*). Finally, a 3D high resolution sequence (VIBE) in the axial plane was obtained for each of the speakers, without phonation, to allow the extraction and co-registration of the mandible and dental casts.

MRI data included the lateral-alveolar /l/ in different word positions: word-initial (e.g., *laca*, *hairspray*), intervocalic (e.g., *sala*, *room*) and word final position (e.g., *sal*, *salt*) in the context of the three EP cardinal vowels ([i,a,u]). For the lateral /L/, only intervocalic context was considered (e.g., *palha*, *straw*).

The subjects that acquired MRI data were seven EP speakers: three female (CO,ER, MC) and four male (LC, JH, JPM, AS). Six of the subjects are native mono-lingual EP speakers and one (JH) is a bilingual speaker (EP and Spanish), ages ranging from 21 to 39, with no history of hearing or speech disorders. An MRI screening form and informed consent were obtained before their participation in the study.

2.2. Segmentation Method

As stated before, segmentation of the structures involved in speech production (tongue, velum, vocal tract) is a difficult and complex task, most often performed manually (e.g.,

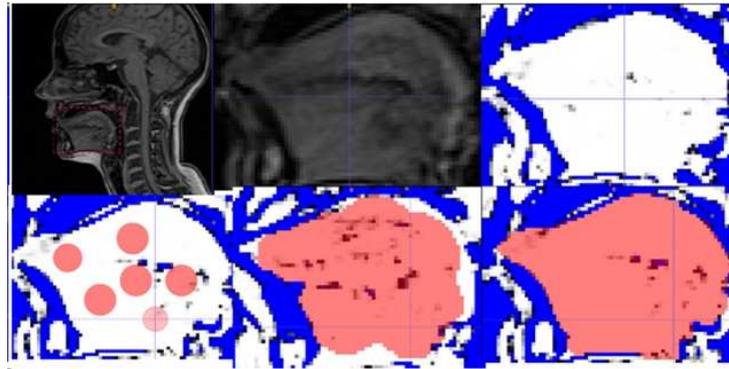


Figure 1: Segmentation steps using ITK SNAP semi-automatic segmentation based on Intensity Regions. From left to right, top: Definition of the region of interest, resampled region, and thresholded image. From left to right, bottom: 3D bubbles placed inside of the tongue, evolution after 121 iterations, and after 182 iterations when the evolution was finished

[BS06]). For lateral sounds these difficulties increase due to the establishment of lateral side branches and of a separate supra-lingual cavity. Although manual segmentation is frequently considered the gold standard, this process is very time consuming, depends on the operator prior knowledge regarding the structures to be segmented, and reliability is a vulnerable issue [PCP00]. Because we are in the presence of a rather extensive database (with seven speakers and several contexts), we have used, whenever possible, semi-automatic tools as they represent a good trade-off between efficiency and accuracy [PCP00].

Segmentation of both the vocal tract and tongue images was based on the general deformable framework. Two open-source tools were used: MevisLab [Mev] and ITK-SNAP toolkit [YPH*06]. Some additional processing, data analysis and visualizations were performed in Matlab. Segmentation of the vocal tract in order to obtain area functions (area variation along the vocal tract) and information regarding the area and the extension of the lateral channels (established around the sides of the tongue when lateral sounds are produced) was more demanding than tongue segmentation. To accomplish this task several steps were needed. All of them were performed with different modules available with Mevislab (e.g. co-registration of the teeth in all the volumes acquired and Curved Multiplanar Reconstruction). From the different segmentation techniques available, stack oriented Live Wire was the method selected for segmentation. More details, on the segmentation of the vocal tract can be found in [MOST10].

In this article, we will focus on tongue segmentation. Tongue segmentation from MR images was performed using ITK-SNAP toolkit. This software application is based on active contours, developed with the main goal of segmenting anatomical structures, in clinical settings, as a reliable alternative to manual segmentation [YPH*06]. ITK-SNAP has mainly been used in the field of neuroimaging to perform segmentations of several intracranial structures from high resolution CT or MRI images (e.g., caudate nucleus, lateral ventricles, hippocampus, brain tumors) with excellent reliability and efficiency [HBG02, HCG03, YPH*06]. Other examples include liver segmentation, bone structures for geometrical modeling (e.g., [LMCP07], [RFL*09]), and also tongue segmentation from MRI images [VWCF09].

ITK-SNAP implements two different automatic methods of segmentation, both based on active contours: Geodesic Active Contours and Region Competition Snakes [HCG03]; [YPH*06]. Manual editing is also possible allowing fully manual segmentations or post processing corrections after semi-automatic segmentations.

In our study, due to the characteristics of our images, the ITK algorithm based on Region Competition Snakes was selected. We have performed a volumetric acquisition, but a compromise between spacial resolution (non-isotropic voxel) and time of acquisition (shorter as possible) had to be established, compromising somewhat the quality of the images (e.g. edges are not clearly defined). Because the tongue is surrounded by air (in oral and pharyngeal cavity) there exists a well defined intensity difference between the object (the tongue) and the background (air). In this scenario it is expected that the algorithm, based on Region Competition Snakes, has a good performance, but a prior resampling and a careful choice of different parameters are needed.

The steps to segment the tongue images include: 1) re-sampling the volume to obtain an isotropic voxel, 2) selecting the region of interest (tongue), 3) defining a threshold, 4) placing seeds or bubbles inside of the tongue, choosing the parameters that control contour evolution and finally running the algorithm (see Figure 1). All of these steps take less than one minute. However, some time was spent previously in the optimization of all the parameters. After that, manual correction can be performed to obtain a more accurate segmentation (particularly at the lateral and inferior sides of the tongue), which can take approximately 5 additional minutes (revise the segmentation provided semi-automatically and remove or add unwanted or missing voxels). Figure 2 presents an example of the segmentation in the three orthogonal planes, and the resulting mesh. When fully manual segmentation (slice by slice) was performed each tongue configuration took about 30 to 40 minutes (depending on the speaker). A mesh of the tongue (which can be smoothed and decimated) was obtained from the segmented volume using ITK-SNAP (internally using the well known marching cubes algorithm [LC87]) allowing a qualitative evaluation (e.g., position, overall tongue configuration, presence and location of grooving). Statistical data concerning the number of voxels and total volume is readily available from ITK statistics.

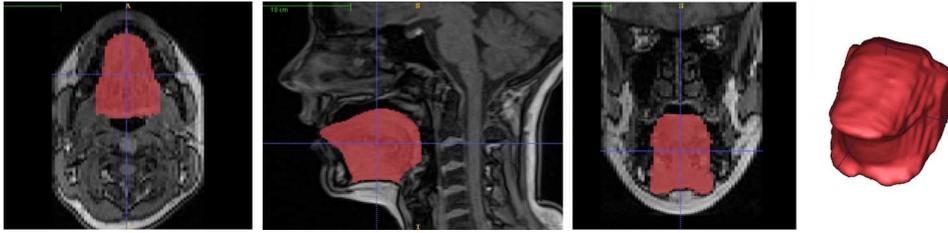


Figure 2: ITK SNAP visualization of a MRI volume obtained for one of the female speakers CO (producing the /l/ as in the word *sul*, south). From left to right: views of the three orthogonal planes (axial, sagittal and coronal) and the mesh obtained after semi-automatic segmentation

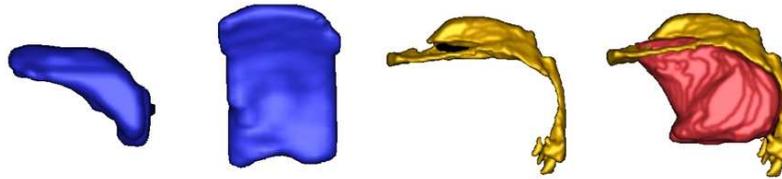


Figure 3: Segmentation of other structures with SNAP. From left to right: lateral and anterior view of the velum, lateral view of vocal tract visualization corresponding to an /l/, and tongue and vocal tract visualizations

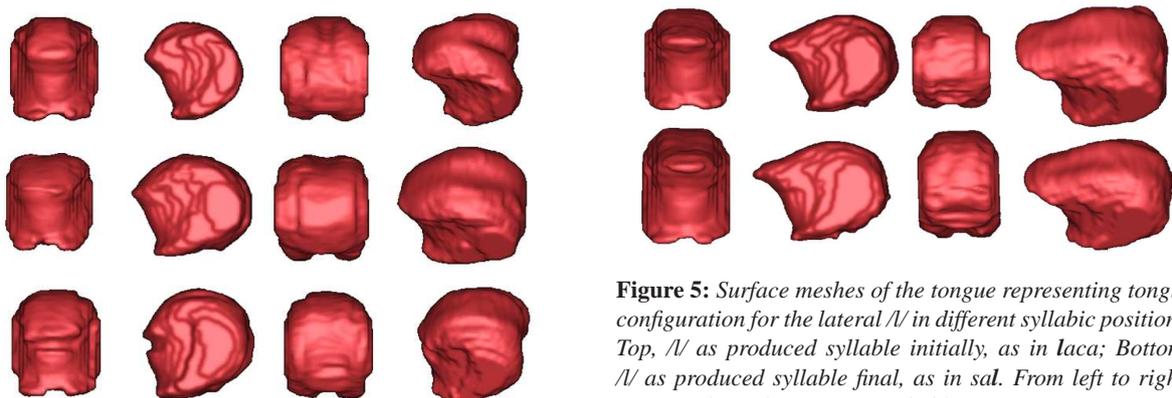


Figure 4: Surface meshes of the tongue representing tongue configuration for the cardinal vowels [i] on the top, [a] on center, and [u] bottom, for the speaker CO. From left to right: anterior, lateral, posterior and oblique views

Figure 5: Surface meshes of the tongue representing tongue configuration for the lateral /l/ in different syllabic positions. Top, /l/ as produced syllable initially, as in *laca*; Bottom: /l/ as produced syllable final, as in *sal*. From left to right: anterior, lateral, posterior and oblique views

The mesh can be exported to other applications, using different formats (e.g. .STL mesh format), offering the possibility of different analysis (e.g., comparing meshes in 3D) or for future modeling purposes, if desired.

We consider that for tongue segmentation the method was efficient, providing a fast method of obtaining 3D visualizations and surface meshes, when compared with manual segmentations. Segmentation of the vocal tract and of the velum was also attempted using a similar approach with ITK-SNAP. For the velum, the results are encouraging but, for the vocal tract, if the target sound is a lateral sound, too much time is needed to correct the output provided semi-automatically. Nevertheless, for a limited number of sounds,

it can provide valuable 3D visualizations, as shown in Figure 3.

2.3. Polygonal Mesh Comparison

Visual analysis and comparison of the polygonal meshes generated from the performed segmentations already provides an important improvement over visual analysis of the segmentations in 2D planes. Nevertheless, comparing the meshes for different sounds (see figure 4 for meshed corresponding to the cardinal vowels [i, a, u]), can still be a tiresome task, and no control is provided over the view conditions (e.g., viewpoint and zoom) to ensure they are the same for all compared meshes. This task can be barely performed if we aimed to compare the same sound (e.g., the /l/), in different syllabic positions or contexts, because they exhibit small differences (see Figure 5).

There are several tools described in the literature

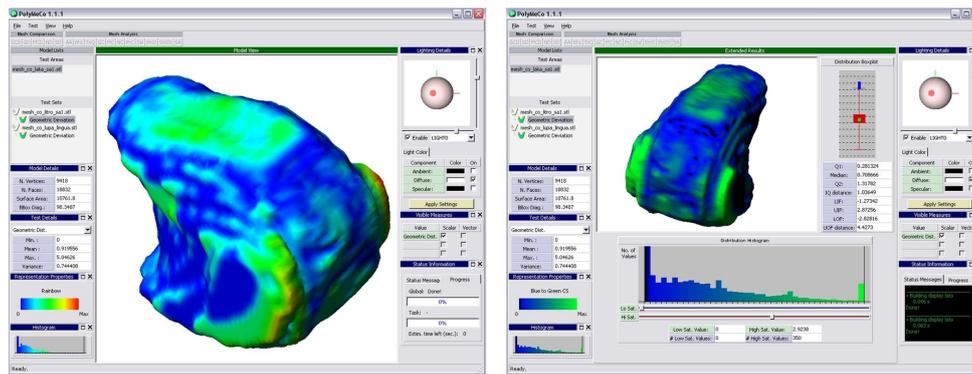


Figure 6: PolyMeCo's graphical user interface. Left, default viewmode showing a mesh colored according to comparison data; right, viewmode showing a colored mesh and comparison data histogram.

such as Metro [CRS98], Meshdev [RFT04] and PolyMeCo [SMSS09] which can be used for polygonal mesh comparison. PolyMeCo was chosen since it provides a user friendly environment where multiple mesh comparisons can be performed and analyzed side-by-side. This can be accomplished using several comparison data representations, such as histograms and colored meshes, while keeping synchronized view conditions (viewpoint, zoom, etc.) among all meshes. Figure 6 shows a snapshot of PolyMeCo's graphical user interface. To analyze the comparison data shown in this article it is important to understand the basics on how this comparison is performed and how it is represented using a colored mesh. When comparing two meshes one is considered the reference to which the other mesh is compared. For each vertex in the reference mesh a corresponding comparison point is found in the surface of the second mesh (which can be a vertex or a surface point, obtained by vertex interpolation), and a metric is used to compute their difference. PolyMeCo provides several metrics for comparing meshes using different criteria such as color, curvature and geometric distance. After computing the selected metric for all vertices in the reference mesh (towards their corresponding surface points in the second mesh) the obtained data can be represented by associating each computed value to the respective reference mesh vertex and presenting the reference mesh with each vertex colored accordingly. Therefore, the comparison data is always represented using the reference mesh. For the comparisons performed in this article the geometric distance (GD) was considered the most adequate metric since it allows the assessment of the differences in size/shape between meshes, as intended.

3. Results

In this section, a detailed articulatory description and a comprehensive phonetic interpretation of the findings is not provided as it is not within the scope of this article. Instead, examples of the comparisons that can be performed, are presented and discussed to highlight possible applications in the speech production field (see figures 7 to 10).

Notice that in each figure all colored models use a common color mapping to allow comparison, i.e., the same color scale representing the same range of values.

Figure 4 presents meshes of the tongue obtained during

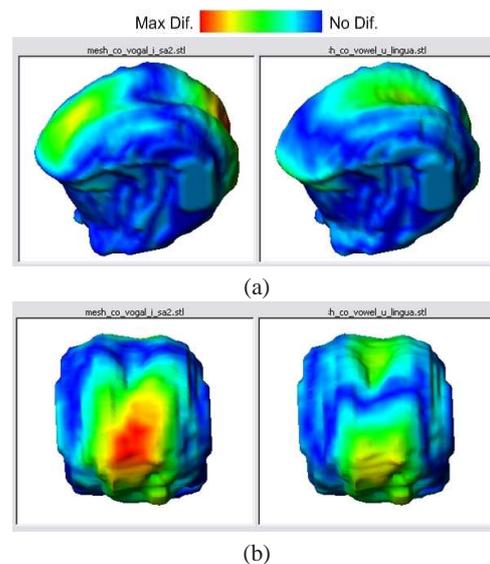


Figure 7: Mesh comparison between the three EP cardinal vowels as produced by the speaker (CO). The mesh corresponding to the vowel [a] was the reference mesh. a) oblique views; b) posterior views. Left: comparison between [a] and [i]; Right: comparison between [a] and [u].

the production of the EP cardinal vowels. These 3D visualizations allow a description of the tongue configuration for each of the vowels. If a careful analyzes of all the views was performed, differences between the sounds will be detected. However, if instead of vowels, similar views of the /l/, in different syllabic positions, had to be considered greater difficulties will arise (see figure 5). Superimposition and comparison of the meshes using a colored model (as shown in Figures 7 to 10) allow the observer to easily detect differences and to have a qualitative measure of the amount of these differences.

3.1. Vowel comparison

The first example presented allows comparison between the three EP cardinal vowels. It has been chosen because the differences between the cardinal vowels are more pronounced and prior knowledge concerning its articulatory description

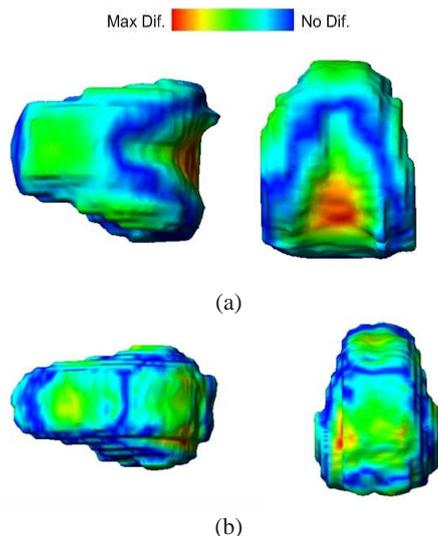


Figure 8: a) Mesh comparison between the vowel [i] (reference mesh) and lateral sound [l] when produced syllable initially followed by the vowel [i]. b) Mesh comparison between the /l/ in different syllabic positions as in *litro* and *til*.

is available. Briefly, concerning the position of the tongue in the oral cavity, the vowels are usually classified with respect to: 1) Tongue Dorsum (TD) height (high and low vowels, and 2) Tongue Backness (anterior and posterior vowels). The [a] is classified (for EP) as a low back vowel, the [i] and [u] are both high vowels, being the latter posterior and the former anterior. Accordingly, we would expect differences between the [i] and [a] at TD level due to differences in TD height, and also at tongue root level due to differences in the degree of backness. In the comparison performed between the [a] and [i] (see Figure 7, left), these differences are clearly demonstrated (red color at the back of the tongue, in the posterior view). Differences in TD height are observed in the oblique view (yellow color). Although, the differences are highest in the midline (probably due to a different degree of grooving) they spread through the sides. When comparing [a] and [u] differences are also detected in TD height and at the back of the tongue although less prominent than those observed between [a] and [i].

3.2. Vowels and consonants

Figure 8 (a) presents comparison between the vowel [i] and the consonant /l/, in the context of [i]. Main differences between these two sounds occur at tongue dorsum, tongue back and lateral sides of the tongue. These differences can be easily justified: the gestures involved in the production of the vowel [i] and the /l/ are antagonics, i.e. in the production of the [i] the tongue is high and forward, for the /l/ the tongue is retracted, somewhat lowered in the middle part of the TD, and the tip is raised to establish contact with the alveolar area. Moreover, due to production requirements, in the /l/ the tongue is compressed towards the midline (accounting for the differences in the lateral sides of the tongue). In Figure 8 (b), the /l/ in two word positions (*litro*, *til*) is compared. Differences observed on the back of the tongue can be attributed to a more retracted tongue in the production of the /l/, in syllable final position.

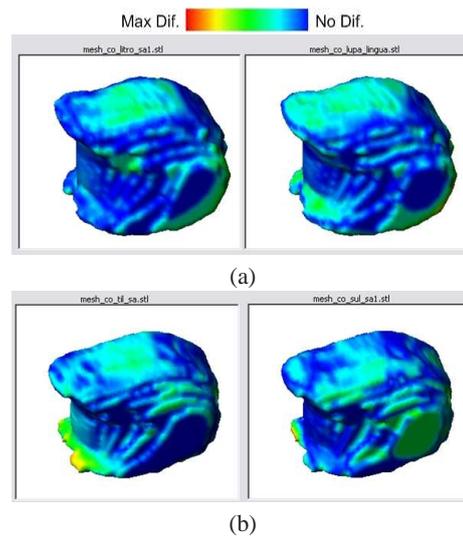


Figure 9: Multiple comparison to evaluate contextual effects (coarticulation). a) Mesh comparison between /l/ in *laca* (reference mesh) with *litro* (left) and *lupa* (right); b) Mesh comparison between the /l/ in *sal* (reference mesh) with *til* (left) and *sul* (right).

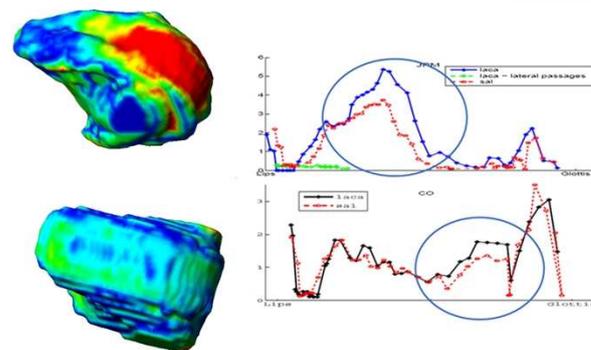


Figure 10: Comparison between speakers. Top (left): comparison of the /l/ as produced in *laca* and *sal* for the speaker JPM, and at right corresponding area functions for the same sounds. Bottom: the same comparison for the speaker CO; common color mapping was used.

3.3. Contextual effects

The extent of lingual coarticulatory effects (i.e. the influence of the vowel in the production of the consonant) could also be qualitatively evaluated by superimposition of the meshes representing the /l/ in the context of the three vowels as produced in the words *laca*, *litro* and *lupa*, and in the words *sal*, *til*, *sul*, as shown in figure 9. As can be observed, the effect of the vowel in the production of the consonant is not prominent, and this effect is even lower when the /l/ is produced syllable final.

3.4. Inter-speaker variability

Furthermore, mesh comparison can also be used to compare and quantify differences between speakers (interspeaker variability) in the production of the same sounds. Figure 10, shows mesh superimposition of /l/ as in the words *laca* and *sal* for each of the two speakers. As can be easily observed differences between both /l/ productions are higher (extensive red color) for the speaker JPM, than for the speaker CO.

Moreover, the information obtained, from tongue comparison, is in agreement with the pattern observed for vocal tract areas.

4. Conclusions

In this study, we used a polygonal mesh comparison tool to compare tongue surface meshes derived from MRI volumes acquired during the production of some EP sounds, uttered by different subjects. Comparison between different sounds, the same sound produced in different contexts and syllabic positions, and comparison between speakers have been presented. To the best of our knowledge no other study attempted this approach in order to compare 3D tongue shapes.

This is a qualitative comparison and the main interest, at the moment, is on where the differences occur and not particularly on the precise amount of those differences. Nevertheless, this comparison already provides very interesting results allowing a better understanding of the articulatory characteristics of the different sounds. In fact, the presented results are already useful, for example, to illustrate different aspects of speech production, in speech related courses.

However, if coarticulatory and/or syllabic effects are to be further evaluated it will be important to quantify these effects. The values obtained for the geometric distance (GD), although representing a relative measure, are a first approach to this quantification. Careful analyzes of these values should be performed and the error introduced in the segmentation procedure should be quantified and accounted for. In a preliminary assessment of the semi-automatic segmentation method, for example, it was observed that for the same speaker the segmented tongue volume is very similar over different sounds, as expected. Nevertheless, using only the total volume is not enough as two voxel sets can have the same volume and be completely different. At this moment, formal validation of the segmentations is being carried out, comparing both manual and semi-automatic techniques and assessing intra- and inter-observer reliability using volume comparison metrics such as the Jaccard coefficient or the Yasnoff discrepancy metric [SSSF*11].

5. Acknowledgements

This research was supported by FCT (Portuguese Research Agency) PhD grants of the first author (SFRH/BD/65183/2009) and second author (SFRH/BD/38073/2007) and HERON II Project (PTDC/EEA-PLP/098298/2008). Authors thank to Professor Miguel Castelo-Branco (IBILI Director) and to the speakers involved in the study.

References

- [And99] ANDRADE A.: On /l/ velarization in European Portuguese. In *International Congress of Phonetic Sciences (ICPhS)* (San Francisco, 1999), pp. 543–546.
- [BBS98] BADIN P., BAILLY G., RAYBAUDI M., SEGEBARTH C.: A Three-Dimensional linear articulatory model based on MRI data. In *International Conference on Spoken Language Processing (ICSLP)* (Sydney, Australia, 1998).
- [BK06] BIRKHOLZ P., KRÄÜGER B. J.: Vocal tract model adaptation using magnetic resonance imaging. In *Proc. of the 7th International Seminar on Speech Production (ISSP'06)* (2006), pp. 493–500.
- [BKZ04] BERNSTEIN M., KING L., ZHOU X.: *Handbook of MRI pulse sequences*. Elsevier Academic Press, Burlington, 2004.
- [BS06] BADIN P., SERRURIER A.: Three-dimensional linear modeling of tongue: articulatory data and models. In *7th International Seminar on Speech Production* (2006).
- [CRS98] CIGNONI P., ROCCHINI C., SCOPIGNO R.: Metro: measuring error on simplified surfaces. *Computer Graphics Forum* 17(2) (1998), 167–174.
- [EB99] ENGWALL O., BADIN P.: Collecting and analysing two and three dimensional MRI data for swedish. *TMQ-QPSR* 4 (1999), 1–28.
- [FR99] FARNETANI E., RECASENS D.: *Coarticulation models in recent speech production theories*. Coarticulation: theory, data and techniques. Cambridge University Press, Cambridge, 1999.
- [FSH*09] FELS S., STAVENESS I., HANNAM A., LLOYD J., ANDERSON P., BATTY C., CHEN H., COMBE C.: Advanced tools for biomechanical modeling of the oral, pharyngeal, and laryngeal complex. In *International symposium on Biomechanics, Healthcare, and Information Science* (2009).
- [GPP06] GÉRARD J., PERRIER P., PAYAN Y.: *3D Biomechanical tongue modeling to study speech production*. Models, Phonetic Processes and techniques. Psychology Press, NY, 2006.
- [HAC*07] HUEBER T., AVERSAMO G., CHOLLER G., DENBY B., DREYFUS G., OUSSAR Y., ROUSSEL P., STONE M.: Eigentongues feature extraction from an Ultrasound-based Silent Speech interface. In *ICASSP* (2007).
- [HBG02] HO S., BULLITT E., GERIG G.: Level-set evolution with Region Competition: Automatic 3-D segmentation of brain tumors. In *Proc. Int. Conf. on Pattern Recognition* (2002).
- [HCG03] HO S., CODY H., GERIG G.: SNAP: A software package for user-guided geodesic snake segmentation. In *Technical Report-university of North Carolina* (2003).
- [LC87] LORENSEN W. E., CLINE H. E.: Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH Comput. Graph.* 21 (August 1987), 163–169.
- [LMCP07] LOPES D., MARTINS J., CAMPOS J., PIRES E.: Modelação geométrica de estruturas humanas baseada em imagens de Tomografia Computorizada. In *CMNE/CILANCE* (Porto, 2007).
- [MCP*08] 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.
- [Mev] MEVISLAB: <http://www.mevislab.de>.
- [MOST10] MARTINS P., OLIVEIRA C., SILVA A., TEIXEIRA A.: Articulatory characteristics of European Portuguese laterals: An 2D and 3D MRI study. In *FALA, 2010* (Vigo, 2010).
- [NAH97] NARAYANAN S., ALWAN A., HAKER K.: 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.
- [Oli09] OLIVEIRA C.: *Do grafema ao gesto: Contributos para o desenvolvimento de um sistema de síntese de base articulatória*. Phd thesis, University of Aveiro, 2009.
- [PCP00] PHAM D., CHENYANG X., PRINCE J.: Current methods in medical imaging segmentation. *Annu. Rev. Biomed. Eng* 2 (2000), 315–337.
- [PPZP03] PERRIER P., PAYAN Y., ZANDIPOUR M., PERKELL J.: Influences of tongue biomechanics on velar-stop consonants: A modeling study. *Journal of Acoustical Society of America* 103 (2003), 1615–1627.
- [RE05] RECASENS D., ESPINOSA A.: 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.
- [RFL*09] RIBEIRO N., FERNANDES D., LOPES D., FOLGADO J., FERNANDES P.: 3-D solid and finite element modeling of biomechanical structures - a software pipeline. In *EUROMECH Solid Mechanics Conference* (Lisboa, 2009).

- [RFT04] ROY M., FOUFOU S., TRUCHETET F.: Mesh comparison using attribute deviation metric. *Internation Journal of Image and Graphics 4(1)* (2004), 1–14.
- [SD02] SHADLE C. H., DAMPER R. I.: Prospects for articulatory synthesis: A position paper. In *4th ISCA Workshop on Speech Synthesis, August, September 2001* (2002), pp. 121–126.
- [SL96] STONE M., LUNDBERG A. J.: Three-dimensional tongue surface shapes of english consonants and vowels. *Journal of the Acoustical Society of America (JASA)* 99, 6 (1996), 3728–3737.
- [SMSS09] SILVA S., MADEIRA J., SOUSA SANTOS B.: Poly-MeCo - an integrated environment for polygonal mesh analysis and comparison. *Computers Graphics 33* (2009), 181–191.
- [SSSF*11] SILVA S., SOUSA SANTOS B., FERREIRA C., MADEIRA J., SILVA A.: A preparatory study to choose similarity metrics for left-ventricle segmentations comparison. In *Proc. SPIE Medical Imaging 2011 (to appear)* (2011).
- [TMS*05] TEIXEIRA A., MARTINEZ R., SILVA L. N., JESUS L. M. T., PRÍNCIPE J. C., VAZ F. A. C.: Simulation of human speech production applied to the study and synthesis of European Portuguese. *EURASIP J. on App. Sig. Proc.* 9 (2005), 1435–1448.
- [TOM*09] TEIXEIRA A., OLIVEIRA C., MARTINS P., DOMINGUES I., SILVA A.: The project HERON. In *I SLTech in Iberian languages* (2009).
- [VWCF09] VOGT F., WILSON C., CHANDER R., FELS S.: Extraction of tongue shapes from magnetic resonance images. In *OPAL* (Vancouver, 2009).
- [YPH*06] YUSHKEVICH P. A., PIVEN J., HAZLETT H. C., SMITH R. G., HO S., GEE J. C., GERIG G.: User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage 31*, 3 (2006), 1116–1128.
- [ZHH08] ZARKOVA N., HEWLETT N., HARDCASTLE W.: An Ultrasound study of lingual coarticulation in children and adults. In *Proc. 8th Int. Seminar on Speech Production* (2008).