International Journal of Computational Vision and Biomechanics
Vol.2 No. 1 (January-June, 2016)

MUK PUBLICATIONS
Open Access Publisher

Morphological Dynamic Imaging of Human Vocal Tract

Sandra Moreira Rua[1] and Diamantino Freitas[2]

[1]*Escola Superior de Tecnologia da Saúde – IPP, Porto, Portugal*

[2]*Faculdade de Engenharia da Universidade do Porto, Portugal*

With the development of high B-value units and ultra-fast acquisition sequences, magnetic resonance imaging (MRI) has been successfully used in speech research to obtain morphological and dynamic data of the vocal tract. This knowledge is useful for the understanding of speech mechanisms and can, in particular. be employed for speech synthesis. The aim of our study was a combination of the morphological and dynamical characterizations of speech, resulting in the production of 3D articulatory models of some relevant Portuguese sounds and syllables. A set of static images in stacks and dynamic sequences were collected during the experiments from subjects that were instructed to vocalize during the acquisition, exploring the capabilities of the equipment. Analysis of the image stacks allowed the extraction of the vocal tract contours and a subsequent non-conventional 3D reconstruction done by combination of orthogonal stacks, allowed visualization and partial measurement of the sampled vocal tract shape.

Keywords: *MRI of Vocal Tract, Magnetic Resonance Imaging, Speech Production, Morphological Study, 3D Vocal Tract Model*

1. INTRODUCTION

The human speech production process has long been a subject of interest, concerning morphological knowledge and speech acoustics, aiming to reach a useful understanding and modelling of all mechanisms involved. The main anatomic aspects and physiology of the vocal tract are common to all people; however speech production is a complex and highly individual mechanism. This recommends that modelling be done with enough accuracy for individual characterization.

Many approaches are available for measuring speech production (for example those based on muscular activity, articulator’s movements and shapes, and signal analysis). Even so the available morphologic and dynamic information is scarce. The first demonstration of 3D reconstructions of the vocal tract shape was carried out by Baer *et al.* (1991), by means of magnetic resonance imaging (MRI). Since then, other speech researchers have used this imaging technique to obtain midsagittal cuts of the vocal tract for the calculation of area functions which are recognized to be of interest for the understanding of the speech production mechanisms (Demolin *et al.* 1995, Demolin *et al.* 1996, Story *et al.* 1996).

Magnetic resonance imaging is a powerful tool for the study of the whole vocal tract with enough safety, through morphological measurements, to get multiplanar and high quality imaging of soft tissue. Several studies of the vocal tract shape, using MRI, have relied on static images (during sustained articulation) as demonstrated for example, by Kroger *et al.* 2000, Serrurier & Badin 2005.

Badin *et al.* (1998) built a 3D statistical articulatory model, based on a set of vocal tract images (in three orientations). The parameters measured are good predictors of 3D shapes for most configurations (20 French sounds sustained during 43 sec.).

In order to assess the influence of sustained sounds (over 30 sec.) required in the majority of MRI studies, Engwall (2000) compared static MRI data with electromagnetic articulography and electropalatography data. The static MRI measurements obtained are a case of very hyperarticulated speech, and sustained sounds are more difficult to hold for the subject. Also, the supine position imposed by the MRI equipment affects the position and shape of the tongue (the most important articulator) causing some obstruction of the pharynx cavity (Engwall 2003).

More rarely than in the case of static, few known dynamic studies have been undertaken to date (Avila-García *et al.* 2004, Demolin *et al.* 2000, Engwall 2003, Engwall 2004, Narayanan *et al.* 2004, Shadle *et al.* 1999), mainly because of imaging device limitations, that make the capture of images of the moving vocal tract difficult. Presently with high B-value units, dynamic analysis becomes more feasible.

More recently, Badin & Serrurier (2006) presented a 3D geometrical database of speech organs for one speaker (a set of 46 sustained French allophones), using linear component analysis of MRI and CT images. This study illustrates the importance of a detailed 3D knowledge of the vocal tract to deal in a more realistic way with speech production and to be useful in the domain of speech rehabilitation (Mády *et al.* 2001, Story 2003).

For the characterization of European Portuguese (EP), only a few studies of nasal vowels have been carried through, at the acoustic production and perceptual levels. These were based on acoustic analysis and electromagnetic articulography (Teixeira *et al.*, 2001; Teixeira *et al.*, 2002; Teixeira *et al.*, 2003). No other MRI study of EP was found up to now.

Therefore, the aims of our study were the construction of 3D models of the vocal tract from MRI data, of some relevant sustained articulations in Portuguese (morphologic study), and the demonstration of an MRI technique to achieve the collection of useful image sequences during speech (dynamic study). We present a new approach for 3D modelling by means of the combination of orthogonal stacks, to describe the vocal tract shape in different articulatory positions.

For the production and modulation of speech, a small number of organs (vocal tract) work together synchronously, however it is well known that there is a lot of variability in this process, the same subject being unable to produce the same sound in the same way many times (Faria *et al.* 1996). Moreover, there is a lot of inter-speaker variability (Nieto-Castanon et al. 1999). In this particular study only the articulation and resonance phase of the human speech production is addressed, because of its relevance for speech research and MRI application.

The remaining of this paper is organized in three sections. The methodology section describes the equipment, corpus and subjects, and the procedures used for the speech study, namely for morphologic and dynamic imaging of the vocal tract. The results are presented in next section, through the exhibition of some magnetic resonance images and some three-dimensional models of the vocal tract. Finally, we present the discussion and conclusions of this work.

2. METHODOLOGY

2.1. The Equipment

Experiments for image acquisition were performed on a Siemens Magneton Symphony 1.5T system, with subjects lying in supine position. A head array coil was used. This study consisted firstly of a static phase (articulations sustained during acquisition), and secondly of a dynamic one (articulations synchronized with acquisition sequences).

2.2. The Corpus and the Subjects

Initially, for the static study, two speech skilled subjects, one male and one female, were involved. The corpus consisted of a set of images collected during sustained articulations of twenty-five Portuguese sounds, oral and nasal vowels, and some consonants. For the dynamic study, four subjects produced several repetitions of sequences of three consonant-vowel syllables (/tu/, /ma/, /pa/) during the acquisition.

The corpus intends to cover the maximum possible range of the articulator's positions, namely, the tongue's, jaw's, lips' and velum's, in order to explore a significant part of the articulatory space.

2.3. Static Study

Initially, a single midsagittal T1-weighted image was recorded with subjects instructed to rest with mouth close and the tongue in full contact with the teeth. This reference image was used for teeth space identification (Figure 1), and later for contour extraction in the segmentation phase of the vocal tract (discussed below).

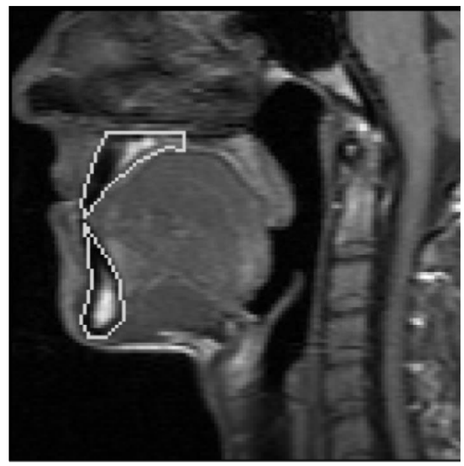

Figure 1: Reference midsagittal image used for teeth identification and posterior contour extraction

Our first approach consisted in the collection of three sets of images of the vocal tract, usually called stacks, to describe the object and represent the sampling of the anatomic volume by means of regularly spaced slices along three orientations. Stacks of sagittal, coronal and transversal MR, T1-weighted images, were recorded using Turbo Spin Echo sequences with a 150mm field of view (FOV). Three sagittal, four coronal and four transversal slices were recorded. The first slices, with 5mm thickness, were recorded during 9 seconds, the coronal and the transversal, both with 6mm thickness and 16mm spaced, were recorded during 9.9 seconds each. An acoustic recording of the produced speech was tried but unfortunately wasn't successful because of the high noise intensity produced in the room during the MR acquisition.

To achieve a correct 3D reconstruction and combination between stacks in the different orientations the following aspects were taken into account, for the sequences programming: (1) consistent use of the same image matrix; (2) no interpolation between slices; (3) similar FOV and no reduction in phase codification; and (4) equal reference coordinates center for all slices. Due to some problems related with segmentation and final alignment of the orthogonal stacks, the transversal stacks weren't incorporated in the reconstruction phase of the study.

2.4. Dynamic Study

The dynamic study was performed using the same principle of MRI cardiac studies, with the modification of a FLASH 2D sequence using the patient's heart beat as trigger, a 300mm field of view and the acquisition parameters: TR = 60ms and TE = 4.4ms.

The subjects tried to synchronize the utterance of the syllables to their own cardiac rhythm by means of the acoustic monitoring of the simple electrocardiogram (ECG or EKG) through a synchronous sound emission conveyed to the subject by an earphone. This method requires some time consumption for subject preparation by the repetition of the syllable for some minutes before acquisition with and without earphone. The idea was to prepare the subject for the difficulty in listening to the heart beat during the noisy acquisition. Each set of images from a single-slice (sagittal) of 6mm thickness was collected during 12 to 22 seconds, resulting in a variable number of images (4-6 images) that were achieved for each sequence, one per cardiac cycle with a regularly increasing shift in synchrony from the start of the cardiac cycle (Figure 2).

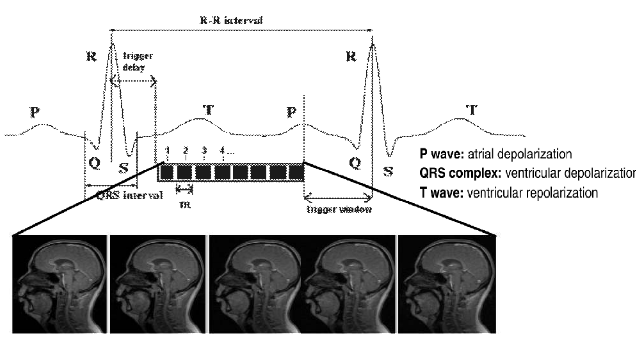

Figure 2: Schematic representation of the EKG monitoring and synchronization technique used for the dynamic MRI sequence acquisition

The EKG is a graphic representation of the electric activity of the heart for tracing the heartbeat (or cardiac cycle). It shows the set of events that occur as the heart contracts, and consists of a P wave, a QRS complex and a T wave. The RR interval is the time duration between two consecutive R waves (an indicator of ventricular rate), and is usually the reference interval for programming slices' acquisitions.

This time shift method of synchronization is in fact an undersampling approach that was necessary because of the important time overhead presented by the actual MR equipment in each single image acquisition. It was based on the assumption that the successive repetition of the utterances, one for each cardiac cycle, was reasonably stationary and therefore a reasonable accuracy was expected in the sampling of different phases from different utterances. This was an available way of achieving a good coverage of the whole articulatory gesture.

2.5. Image Analysis and 3D Vocal Tract Models

Image analysis and 3D model construction was accomplished in two stages, namely: (1) image segmentation using the S*egmenting Assistant*, a 3D editing plugin of *Image J* the image processing software developed by the National Institute of Health (*http:// rsb.info.nih.gov/ij/*) and subsequent 3D reconstruction, and (2) graphic representation and combination of orthogonal stacks using the *Blender* software for 3D graphics creation, version 2.41 (*http://www.blender.org*).

Although all the existing segmentation techniques employed in MRI images segmentation have more or less automated contour extraction tools, this procedure is still hard and complex, due to the characteristics of the images produced by the MRI technique, on one hand, and to the main anatomic features of the vocal tract (i.e. shape length and adjacent organs) on the other hand. Specifically the main anatomic issues are:

- The role of the epiglottis and piriform sinus in speech production is unidentified, consequently their relevance for delimitation and its inclusion in the segmentation process is not clear yet;
- The teeth aren't shown in MRI;
- The end of the vocal tract remains to be defined;
- The similar MRI signal intensities of some organs or structures located in the boundaries of the vocal tract (vertebral column for example) allow automatic contours to "escape" and complicate the segmentation.

Figure 3 depicts an example of the contour extraction problems that occur due to the similarity of pixel intensity values between anatomic structures.

Figure 3: Leakage of contours obtained directly by a *threshold* technique

A few segmentation methods have been described in speech studies, for vocal tract contours extraction from MR images, based on manual edition, such as Bézier curves, and *threshold* contours. Soquet *et al.* (1998) compared different approaches over the same data in order to assess the accuracy of three segmentation methods: manual, *threshold* and elastic. This study shows (1) that the different methods give comparable results with a somewhat lower dispersion for the *threshold* method, and (2) the settings of the parameters of each method (contrast, *threshold* level, free-form curve) have an impact on the resulting area.

The histogram-derived *threshold* technique[1] was chosen because of its semi-automatism, flexibility in contours adjustment, and lesser dependency of the user's influence when compared to manual methods. The image segmentation of the airway from the surrounding tissues for extraction of the contours of the vocal tract was then obtained by the following sequence of procedures:

(a) Identification and closure of the vocal tract area of interest, mandatory closure of the mouth, larynx, vertebral column and velum, through the manual superimposition of opaque objects;

(b) Manual pasting of teeth image (for now only on the sagittal stacks), after extraction of the teeth contours from the initially acquired sagittal anatomic reference image;

(c) Extraction of the contours of the vocal tract, for each image of 2D slice using the *Image J* semi-automatic *threshold* technique.

Figure 4 shows, at left, the result of the manual identification and closure of the region of interest of the vocal tract with solid opaque boxes, and the pasting of the contours of the teeth. The extraction of contours of the vocal tract was then performed by outlining the area of interest, slice by slice. The result can be observed on the right part of the figure.

Each outline is defined by the minimum and maximum levels in the area, moving and controlling histogram intensities values. The boundary is given after adjusting some parameters ("min" and "max" levels of intensities and the "centroid values") to match the outline with the area of interest.

Outlines were subsequently used to generate a skin (a 3D object), after importing the contours in *.shapes* format, into the *Blender* software.

For each articulatory position, the next phase was the combination of sagittal and coronal outlines (2D curves). To make this possible, it is required that the outlines be well aligned – this process is sometimes called image registration. In computational vision, the term image registration means the process of transforming the different sets of images into one coordinate system, what is necessary in order to be able to compare or integrate the data obtained from different measurements.

The image registration was performed using the data contained in the header of the image stacks that belongs to the DICOM protocol (Digital Imaging and Communications in Medicine) for the transmission of medical images. This protocol stores each image's and each stack's full metadata, including the specific attributes that have a positional and magnification interest: image orientation, image position, pixel spacing and slice thickness.

The alignment was obtained by the calculation of the reference pixel location, which is given by the product of the image position attribute values by the image resolution value.

3. RESULTS AND DISCUSSION

The static study was designed to obtain the morphologic data of most of the range of the articulator's positions aiming the imaging characterization of Portuguese sounds. Sagittal data is particularly useful in the study of the whole vocal tract anatomy, demonstrating the main aspects of the shape and positions of some articulators, e.g. tongue, lips and velum (Figure 5).

This figure allows a comparison between images obtained from a male and a female subjects, during the production of identical phonemes. Despite the expected vocal tract anatomic differences a remarkable similarity of movements of the articulators, namely the tongue, can be observed. These differences don't modify the quality of the results because the slices programming was adjusted for each subject and so there is no expected influence on the anatomic coverage of the obtained 3D models.

On the other hand, MRI coronal stacks bring specific lateral information, especially useful in the study of lips and tongue shapes. As an example, the images in figure 6, with some manual editing, show lips rounding (6a) indicated by the filled area, tongue shape with bézier

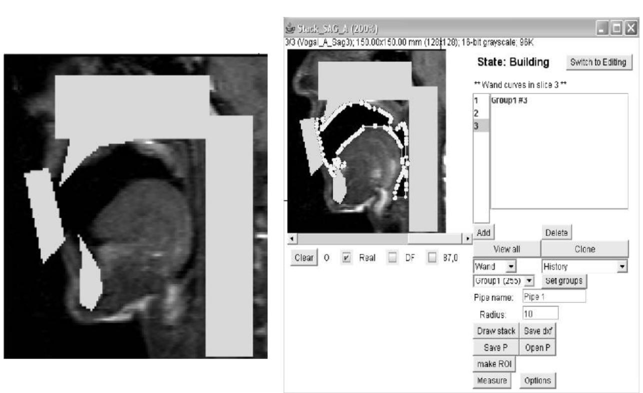

Figure 4: Image segmentation of the airway

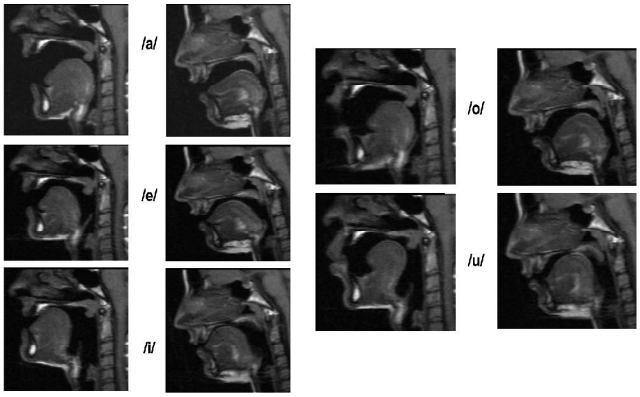

Figure 5: Midsagittal images of Portuguese oral vowels artificially sustained by a male (left) and a female (right)

contour (6c) and the relationship between the palate and the tongue (6b, 6d).

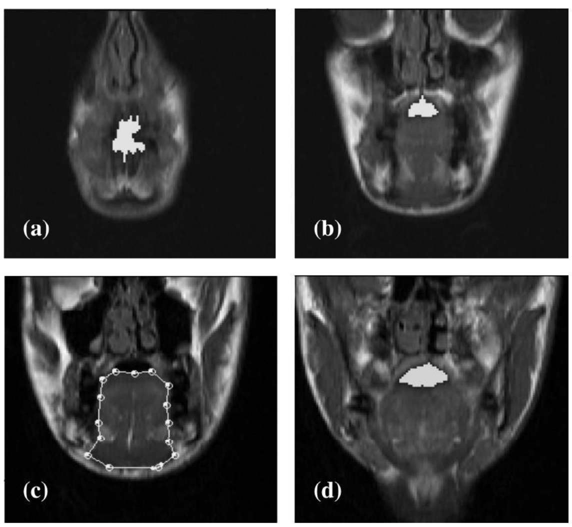

Figure 6: Coronal stack of the Portuguese consonant /ch/

In particular, the coronal data is important for 3D modelling of the vocal tract, because some articulatory situations lead to occlusions in the midsagittal plane, while lateral channels are maintained open (e.g. lateral consonants, nasals vowels).

Nasality is produced by the lowering of the velum, so part of the air stream is free to pass through the nasal cavity, which is known to function as an anti-resonance chamber. By adding this new feature the speech sounds are significantly altered, becoming more complex, as is well known. In Portuguese there is a special interest in nasal sounds due to their frequent use in common speech.

Figure 7 depicts two situations, with some image processing, of the phenomenon of velum lowering (pointed by the white arrow) during the sustained vocalization of two Portuguese nasal vowels and its relationship with the tongue (pointed by the two arrows on the right image). The white filled area corresponds to the vocal tract and the vocal organs are represented by contours.

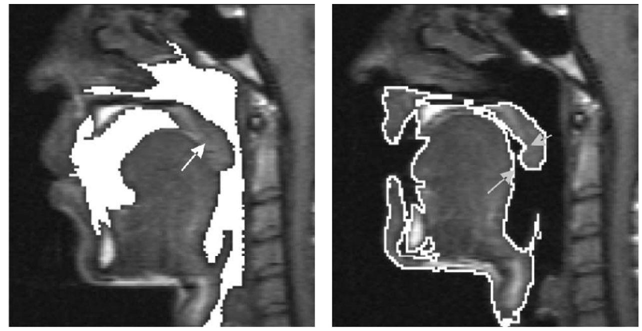

Figure 7: Sagittal images of the vocal tract from a male subject uttering Portuguese nasal vowels: [ũ] (left) and [ĩ] (right)

3.1. Some 3D Vocal Tract Models

The following images (Figure 8) represent different perspectives of the 3D model obtained for the vowel /u/ . The blue skin represents the union of the three outlines extracted from the sagittal stack (8a). The red skin represents the union of the four outlines extracted from the coronal stack (8b). It must be observed that these reconstructions are not yet closed as should be for a whole vocal tract reconstruction. This closing of the skin by unification of the different stacks is the next step in the processing.

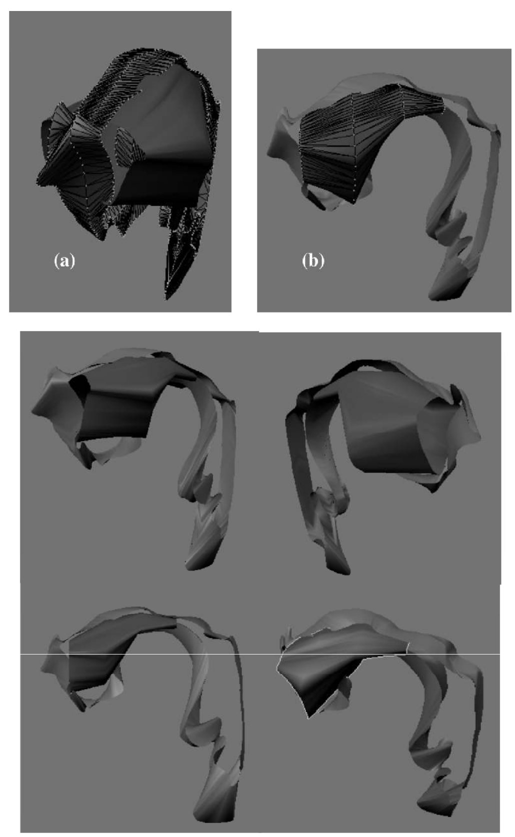

Figure 8: Three-dimensional model of the vowel /u/ uttered by a male subject in different views

For a better understanding of these 3D models, the next figure illustrates one model superimposed to the representation of the head outline with anatomic reference marks and dimensions.

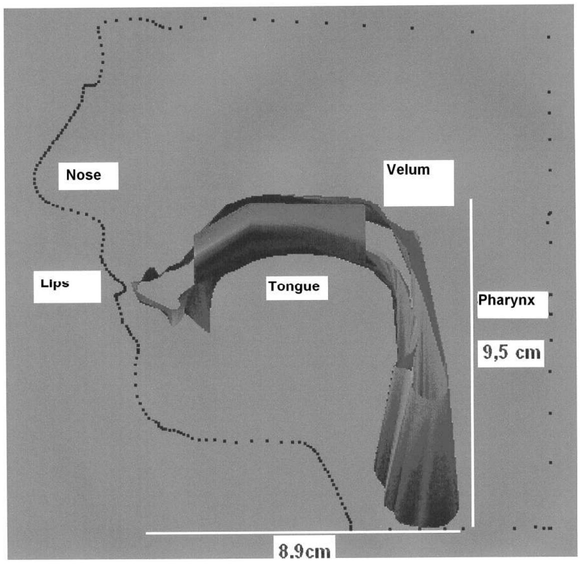

Figure 9: Three-dimensional model of the vowel /a/ referenced to the anatomy with dimensions

Comparatively, we present the 3D model of the same vowel but for a nasal sound production (Figure 10).The different view perspectives presented allow the identification of the velum lowering, and especially the partial closure of the oral cavity.

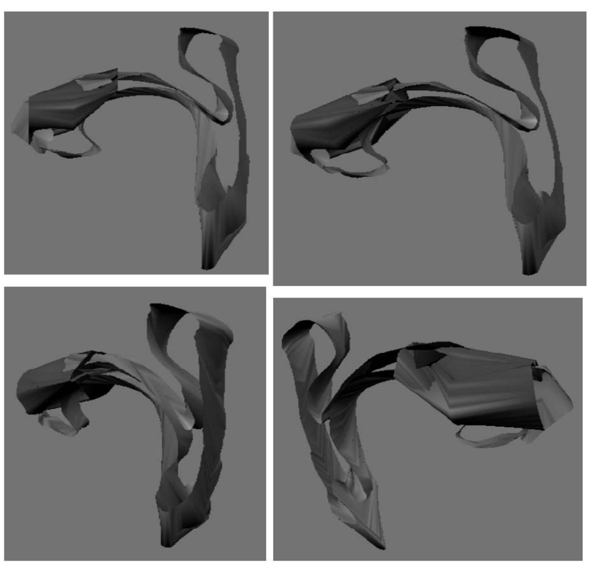

Figure 10: Three-dimensional model of the nasal vowel [ũ] uttered by a male subject

The following models (Figure 11) intend to demonstrate the relevance of coronal stacks for the characterization of lateral consonants of Portuguese. As observed, the coronal data is especially useful in these situations, where tongue is in almost total contact with the palate at central position, and the air escapes by the tongue's side. It should be noted that the removal of the teeth spaces was not done in these models. This has influence in the visible size of the lateral openings, which appear to be much larger in the lower part than in reality.

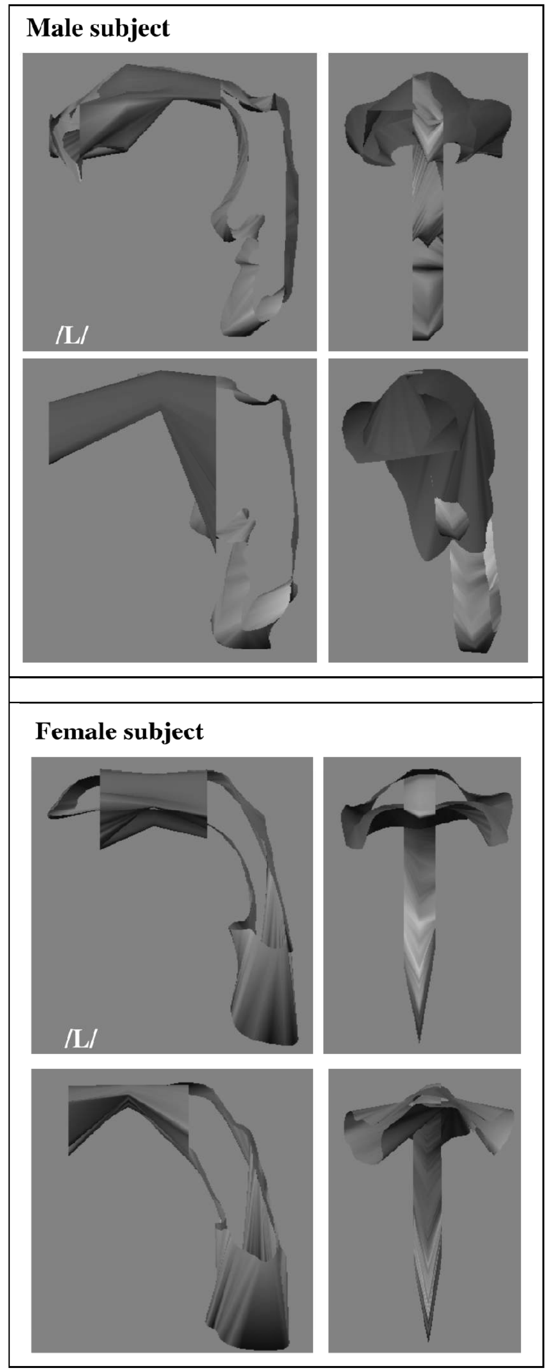

Figure 11: Three-dimensional models of lateral consonants in different perspectives

The 3D presented models demonstrate several essential features needed for the articulatory description of speech, namely:

- shape of the tongue;
- position of the tongue in the oral cavity, and relation to the palate;

- mouth opening degree and shape;
- lateral dimension of oral cavity;
- length of the vocal tract;
- velum position (open or close).

In these 3D models, some differences in vertical length between sagittal and coronal stacks, in some sounds were observed, resulting in some registration errors. This could reflect the specific variability of the speaker in sound production, due in this case, to the fact that acquisitions of different orientations were separately taken (first sagittal images, and subsequently coronal images), to minimize the subject's effort needed for an extra long utterance. On the other hand, the segmentation process by the *threshold* technique has some implications for the determination of the vocal tract area contour as well.

3.2. Dynamic Imaging and Shape Extraction

A variable number of images (sagittal slices) were obtained by dynamic MRI, according to the cardiac cycle (heart frequency) of each subject, followed by the assembly of all images for a movie presentation. The best image features were obtained for the syllable /tu/, because the articulatory positions are very different between single sounds, compared with the other syllables.

Figure 11 represents one of the image sequences, of a male subject, during the repetition of the syllable /tu/, by a male subject, which is partially represented by some contours (using the *threshold* technique). The successive vocal tract shapes acquired present quite small differences that may pass undetected when examined separately. In order to highlight the quantity and quality of the incremental modifications between successive images a set of contours were extracted and placed side by side. The differences are then quite well viewable. This illustration of the movement is therefore more observable as temporal modifications of the articulator's positions.

The sequence of images in Figure 11 depict the consonant /t/, which is characterized by a sequence starting from lips closure (11a) and then followed by the approximation of the tip of the tongue to the upper alveolar region (11b). After that, the tongue moves back with the opening of the mouth for the production of the vowel /u/ (11c), and starts the inverse process for the next repetition (11d).

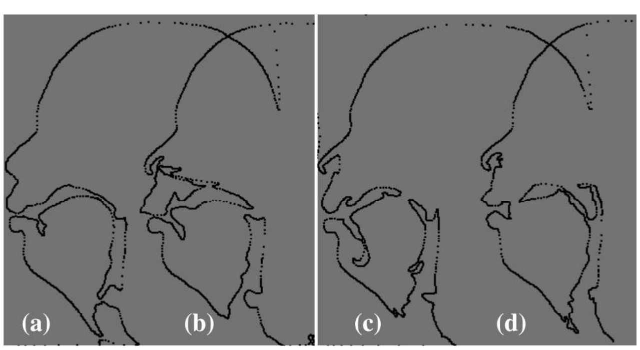

Figure 12: Contours extracted from midsagittal images obtained in the dynamic study by the repetition of the sequence /tu/

When considering the differences amongst the subjects, by image comparison, these dynamic studies demonstrate the actual variability in sounds production between subjects, not only due to anatomic differences, but also because each subject uses different strategies for motion control and articulation.

4. DISCUSSION AND CONCLUSION

Because of the anatomic and dynamic complexity of the vocal tract, the study of speech production has a multidisciplinary interest, carried out by several researches employing different techniques; despite some limitations (concerning time acquisition, number of equipments available, etc.), the use and progress of MRI has been demonstrated to be important in the characterization not only of the anatomic shape but also of the articulation.

The accurate measurement of the vocal tract shape during phonation is otherwise the most important part of speech production studies aiming speech articulatory modelling. For a very long time, the midsagittal plane was used to measure the vocal tract (and help to obtain area functions from these data). This approach, according to Badin *et al.* (2000), leads to a number of problems: (1) the need for a model converting midsagittal contours to area functions, (2) the difficulty of modelling lateral consonants and (3) the limitation of acoustical simulations to the plane wave mode only.

Therefore, three-dimensional data is obviously needed in order to get a more realistic representation that may lead to better models. Demolin *et al.* (1996) proposed a new MRI technique to make measurements of sagittal, coronal, coronal oblique and transversal planes; the extracted areas were placed on a flux line of the vocal tract from a midsagittal slice. These data offer, according to the authors, promising perspectives for the realization of true 3D measures of the vocal tract. Currently, with the increasing availability of MRI devices and powerful image processing means it is more and more feasible to obtain these 3D vocal tract data with a reasonable acquisition speed.

In our study, a considerable number and diversity of images has been collected aiming at not only morphological but also a dynamic characterization, by exploring various MRI techniques. We tried to collect a number of images with enough anatomic resolution, maximum vocal tract extension of representative speech gestures, minimizing speaker effort (reducing hyperarticulation). The image material was analysed and processed resulting in the construction of 3D models for the entire corpus (3D geometrical database). For almost all 3D models obtained for Portuguese sounds the morphologic data shows that both orientations slices (sagittal and coronal) are useful for the knowledge of the vocal tract shape during speech production. Articulator's positions are better demonstrated in sagittal images, and the coronal images allow the observation of the lateral dimension of oral cavity.

From this study important problems have to be taken into account:

(1) Difficulties were encountered in the contours extraction, due to the non-visualization of the teeth by MRI, especially important in coronal images;
(2) Limitation on the number of images collected for a given acquisition time, due to the compromise between time and resolution and the equipment's protocols, and subject effort;
(3) Dynamic acquisition time was limited by time factors namely by the image acquisition trigger technique and several functions of the equipment (e.g., time required to activate and deactivate field gradients);
(4) Difficulties in acoustical recording because of the high acoustical noise level in the acquisition room.

Some dimensional errors have been verified in the combination of stacks, which we think may be related to subject variability in speech production and/or on segmentation problems. The set of images for each orientation was achieved separately to minimize subject effort during sustained articulation, and due to the large range of sounds studied.

In brief, these collected images permit in theory the determination of areas useful for articulatory studies, including speech synthesis. The morphologic images give a better understanding about sounds production, about inter-speakers speech variability, and subject's comparisons. The description of the movements of articulation involved in normal speech that was observed by the dynamic study, allows not only a better knowledge of all mechanisms involved, but also may contribute for understanding speech co-articulation phenomena.

The use of MRI can provide very useful and precise morphological information about the position and shape of the different articulators involved in speech production and also of their dynamics, although, in the present work, still with some speed restrictions due to the limited temporal resolution of present day available equipment and protocols. Improvement and refinement of the protocols is also necessary to take still more advantage of the capabilities of the equipment. The development of a speech-specific synchronization technique and a trigger device MRI-compatible is need for a better improvement of dynamic studies. New potentials are expected with the application of this methodology in 3T MRI equipment.

MRI limitations are mainly: the current impossibility of acoustic recording (because of the noise produced), the required corporal posture of the subjects (supine position), and the non-identification of the teeth. Meanwhile, these can be partially compensated by merging these data with other techniques, for example electropalatography, ultrasound or electromagnetic articulography. These combinations will be done in the near future in the sequence of the present research.

The image segmentation process is quite hard and time consuming, so the choice of segmentation tool is important. The *threshold* technique is the more useful in boundaries extraction, making models more realistic, by comparing with manual segmentation. However segmentation can be improved by the development of a device or technique for teeth identification, without the need to use other methods, e.g. computed tomography.

The completion of the construction of the skins for the hybrid models made from sagittal and coronal stacks is the next step in the way to obtain a complete 3D anatomical model of the vocal tract, prepared for the subsequent prediction of the acoustic output. The extension of the dynamic sequences obtained to other sequences is also important in terms of coverage of the study, and will be done in the near future.

The obtained data can be used for several studies in articulatory phonetics and in vocal tract modelling for speech synthesis, with applications to speech pathology, linguistics, and artificial speech. The approach presented in this article can be applied to any language, and may serve as a support and education means to speech therapists.

ACKNOWLEDGMENT

Images were acquired at the Radiology Department of the Hospital S. João, Porto, with the collaboration of Isabel Ramos (Professor of Faculdade de Medicina da Universidade do Porto and Department Director) and the technical staff, which is gratefully acknowledged.

NOTE

1. The threshold is chosen from the brightness histogram of the region that we wish to segment.

REFERENCES

[1] Avila-García, M. S., Carter, J. N., Damper, R. I. (2004), Extracting Tongue Shape Dynamics from Magnetic Resonance Image Sequences. *Transactions on Engineering, Computing and Technology V2, December*, ISSN, 288-291.

[2] Badin P., Bailly G., Raybaudi M., Segebarth C. (1998), A three-dimensional linear articulatory model based on MRI data.

Proceed. 3rd ESCA / COCOSDA International Workshop on Speech Synthesis, Jenolan Caves, Australia, 249-254.

[3] Badin, P., Borel, P., Bailly, G., Revéret, L., Baciu, M., Segebarth, C. 2000.Towards an audiovisual virtual talking head: 3D articulatory modeling of tongue, lips and face based on MRI and video images. *Proceed. 5th Speech Production Seminar*, München, Germany, 261-264.

[4] Badin, P., Serrurier, A. 2006. Three-dimensional Modeling of Speech Organs: Articulatory Data and Models, *IEICE Technical Committee on Speech, Kanazawa*, Japan, 29-34.

[5] Demolin D., Hombert J. M., Lecuit V., Segebarth C., Soquet A. (1995), A Magnetic Resonance Imaging study of French vowels. *Proceed. 4th EuroSpeech 95 Conference*, Madrid, Spain, 2235-2238.

[6] Demolin D., Metens T. , Soquet A. (1996), Three-dimensional Measurement of the Vocal Tract by MRI. Pro*c. 4th International Conference on Spoken Language Processing (ICSLP 9*6), Philadelphia, USA, 272-275.

[7] Demolin, D., Metens, T., Soquet, A. (2000), Real time MRI and articulatory coordinations in vowels. *In Proc. 5 th Speech Production Seminar*. München, Germany.

[8] Engwall, O. (2000), Are static MRI representative of dynamic speech? Results from a comparative study using MRI, EPG and EMA. *Proceed. of 6th International Conference on Spoken Language Processing* (*ICSLP*), Beijing, China, 17-20.

[9] Engwall, O. (2003), A revisit to the Application of MRI to the Analysis of Speech Production - Testing our assumptions. *Proceed. 6th Int Seminar on Speech Production*. Sydney.

[10] Engwall, O. (2004), From real-time MRI to 3D tongue movements. *Proc of ICSLP 2004*, vol. II, October, Jeju Island: Korea, *1109-1112*.

[11] Faria, I. H., Pedro, E. R., Duarte, I., Gouveia, C. A. M. *Introdução à Linguística Geral e Portuguesa*. Lisboa: Caminho, 1996.

[12] Kröger, B. J., Winkler, R., Mooshammer, C., Pompino-Marschall, B. (2000), Estimation of Vocal Tract Area Function from Magnetic Resonance Imaging: Preliminary Results. *Proceed. 5th Seminar On Speech Production: Models And Data, May 2000*. München: Germany.

[13] Mády, K., Sader, R., Zimmermann, A., Hoole, P., Beer, A., Zeilhofe, H., Hannig, C. (2001), Use of real-time MRI in assessment of consonant articulation before and after tongue surgery and tongue reconstruction. *Proceed. 4th International Speech Motor Conference*, Nijmegen, Netherlands, 142-145.

[14] Narayanan, S., Nayak, K., Lee, S., Sethy, A., Byrd, D. (2004), An Approach to Real-time Magnetic Resonance Imaging for Speech Production. *Journal Acoustical Society of America* 115(4): 1771-76.

[15] Nieto-Castanon, A., Guenther, F. H. (1999), Constructing Speaker-Specific Articulatory Vocal Tract Models for Testing Speech Motor Control Hypotheses. ICPH, San Francisco, 2271-74.

[16] Serrurier, A. & Badin, P. (2005), A Three-dimensional Linear Articulatory Model of Velum based on MRI data. *Interspeech 2005: Eurospeech, 9th European Conference on Speech Communication and Technology*, Lisbon, Portugal, 2161-2164.

[17] Shadle, C. H., Mohammad, M., Carter, J. N., Jackson, P. J. B. (1999), Multi-planar dynamic magnetic resonance imaging: new tools for speech research. *Proceed. 14th International Congress of Phonetic Sciences (ICPhS 99)*. S. Francisco: USA.

[18] Soquet, A., Lecuit, V., Metens, T., Nazarian, B., Demolin, D. (1998), Segmentation of the Airway from the Surrounding Tissues on Magnetic Resonance Images: A comparative study. *Proceed. ICSLP*. Sydney, 3083-3086.

[19] Story B., Titze I., Hoffman E. (1996), Vocal tract area functions from Magnetic Resonance Imaging. *J. Acoust. Soc. Am. (JASA)*, 100(1), 537-554.

[20] Story, B. H. (2003), Using imaging and modeling techniques to understand the relation between vocal tract shape and acoustic characteristics, *Proceed. of the Stockholm Music Acoustics Conference*.

[21] Teixeira, A. *et al*. (2002), SAPWindows – Towards a Versatile Modular Articulatory Synthesizer, *Proceedings of 2002 IEEE Workshop onSpeech Synthesis*. Aveiro, Portugal, 31-34.

[22] Teixeira, A., Moutinho, L. C., Coimbra, R. L. (2003), Production, Acoustic and Perceptual Studies on European Portuguese Vowels Height, *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, CD-Rom (ISBN: 1-876346-48-5), 3033-3036.

[23] Teixeira, A., Vaz, F. (2001), European Portuguese Nasal Vowels: An EMMA Study. *Proc. Eurospeech 2001*. Institut EETA, Aveiro, Portugal, 1483-1486.