Towards Dynamic Magnetic Resonance Imaging of the Vocal Tract during Speech Production

Sandra M. Rua Ventura

Radiology, School of Allied Health Science – IPP

R. Valente Perfeito 322, 4400-330 Vila Nova de Gaia, Portugal

Phone : +351 22 206 1000, fax: +351 22 206 1001, e-mail: smr@estsp.ipp.pt

Diamantino Rui S. Freitas

Department of Electrical Engineering and Computers

Faculty of Engineering, University of Porto

Rua Dr. Roberto Frias, s/n 4200-465 Porto, Portugal

Phone: +351 22 508 1837, fax: +351 22 508 1440, e-mail: dfreitas@fe.up.pt

João Manuel R. S. Tavares[1]

Department of Mechanical Engineering, Faculty of Engineering, University of Porto

Laboratory of Optics and Experimental Mechanics, Institute of Mechanical Engineering and Industrial Management

Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal

Phone: +351 22 508 1487, fax: +351 22 508 1445, e-mail: tavares@fe.up.pt

---

[1] Corresponding author

Towards Dynamic Magnetic Resonance Imaging of the Vocal Tract during Speech Production

Abstract

The most recent and significant magnetic resonance imaging (MRI) improvements allow for the visualization of the vocal tract during speech production, which has revealed to be a powerful tool in dynamic speech research. However, a synchronization technique with enhanced temporal resolution is still required.

Objectives/ Hypothesis and methods: Throughout this work, a technique for the dynamic study of the vocal tract with MRI by using the heart's signal to synchronize and trigger the imaging acquisition process is presented and described. The technique in question is then used in the measurement of four speech articulatory parameters in order to assess three different syllables (articulatory gestures) of European Portuguese Language.

Study design: Transversal.

Results: The acquired MR images are automatically reconstructed so as to result in a variable sequence of images (slices) of different vocal tract shapes in articulatory positions associated with Portuguese speech sounds.

Conclusions: The knowledge obtained as a result of the proposed technique represents a direct contribution to the improvement of speech synthesis algorithms, thereby allowing for novel perceptions in coarticulation studies, in addition to providing further efficient clinical guidelines in the pursuit of more proficient speech rehabilitation processes.

Keywords: Dynamic Study; Vocal Tract Shape; Synchronization Technique; Articulators' Movements; Portuguese Speech Sounds

1. Introduction

The initial application of Magnetic Resonance Imaging (MRI) presented by Baer et al. [1] constituted a new challenge in speech production research, particularly as far as vocal tract visualization is concerned. Initially, most of the studies presented essentially relied on MRI acquisitions of vowel production [2, 3], followed by consonants [4, 5], for different speech languages, such as French [6, 7], German [8, 9] and Japanese [10, 11].

Several approaches have been proposed for the study of the vocal tract based on MRI ranging from static to dynamic studies. Due to the most recent improvements in MRI systems, which have allowed for the visualization of the movements of the vocal tract organs during speech production, some recent works/studies have been produced in the area of real-time imaging acquisition [12, 13]. MRI allows for the acquisition of multiplanar images (slices) with a relatively high image quality and resolution of soft tissues, without producing any harmful effects on patients. As such, it has been considered to be one of the most promising tools for vocal tract studies [14-17].

Due to the fact that speech production is a complex and intrasubject mechanism, most of the reviewed studies are rather limited; this because many anatomical and physiological aspects of the vocal tract are different amongst the subjects who participated in the study, thus resulting in considerable individual and inter-speaker variability in the process. This can be exemplified by the fact that a subject is usually unable to repeatedly produce a particular sound in an identical manner [18, 19].

The European Portuguese (EP) speech language is one of the richest in terms of vowel phonologies and nasals sounds. Despite this, knowledge of the EP speech language remains scarce, namely as far as the speech production mechanism and the resulting coarticulation from magnetic resonance images are concerned. Our approach is thus a pioneer in terms of the dynamic study of EP sounds based on MRI. Additionally, we have developed a protocol for the vocal tract study during speech production based on MRI which was an urgent necessity.

The main objective of our research was the evaluation of the employment of a MRI dynamic technique so as to study the vocal tract during EP speech production. Therefore, throughout this paper, the image sequences of the vocal tract shape in a number of different articulatory positions, which have been attained by using the heart's signal to synchronize and trigger the MRI acquisition process, are presented and discussed. Furthermore, in order to demonstrate the applicability of the approach adopted by us, the assessment of articulatory gestures during the syllable utterance by four articulatory parameters will be dealt with in the present paper.

This paper is organized as follows: in the following section, the specificities of the vocal tract, in addition to several approaches for its dynamic MRI, which have been considered, are presented. Next, the equipment used, the corpus of speech which was studied and the proposed MRI dynamic protocol are described. Following that, the sequence of Cine-MR images for each syllable, which was obtained is disclosed and discussed. In the final section of this paper, a number of conclusions are presented and further suggestions for future work are indicated.

1.1. *Vocal Tract 's Specificities*

The primary function of the human vocal tract is the articulation of sound involving a set of organs, called articulators, which work together in order to produce speech. The vocal tract is similar to an inverted L-shaped tube of a significant length extending from the vocal folds to the lips, in addition to a side branch leading to the nasal cavity. The articulators include the lips, the tongue, the teeth, the palate (the hard and soft palate or velum) and the pharynx which form a set of cavities for the air flow during speech production, Figure 1.

Due to the wide range of movements and flexibility of a number of the vocal tract's organs, in particular, of the tongue, it is difficult to examine and measure their exact movements. Furthermore, a phenomenon known as coarticulation may occur, as a result of the overlapping of adjacent articulators or due to the simultaneous movement of two articulators during different phonemes. This being the case, in [20] some preliminary results required to study coarticulation and articulatory compensations during normal real-time speech production are presented. Subsequently, in [3] midsagittal distances, measured in both static and real-time MR images from the same speaker and sound are compared, thereby demonstrating that the adopted real-time technique provides accurate and reliable information about this phenomena.

1.2. Approaches for Dynamic MRI

Dynamic studies are an important and complex challenge in terms of speech production research, mostly because of the fact that common MRI scanners are limited when it comes to accomplishing a suitable speed/resolution relation. In these studies, kinematics information is based on the acquisition of a gated image sequence with several repetitions of speech in order to reconstruct the impression of articulatory movements over a long period of time. Furthermore, vocal tract movements can be imaged directly in real-time studies. However, high temporal resolution is required to perform motion tracking and analysis in a successful manner.

Previous studies have used distinct techniques in order to image the dynamic vocal tract in various situations, such those based on fast MR imaging sequences, on tagged MRI images or on suitable synchronization methods. The adaptation of an ultrafast MRI sequence to achieve a dynamic continuous monitoring of the vocal tract's movements has been the approach most commonly employed, allowing for resolutions of up to 4 frames/sec [20], 8-9 frames/sec [9, 13], 24 frames/sec [5] and 30 frames/sec [11].

Most MRI systems have the capability to acquire images following an external stimulus (trigger) with an adjustable delay. Considering the fact that feature, as well as making the delay varies continuously, the acquisition and reconstruction of pseudo-moving images are possible in the case of repetitive periodic motion, thus permitting under-sampling. This technique is widely used for cardiac studies and may be explained as follows: The accomplishment of an enhanced feature represents the regularity of the periodic phenomenon. This in turn, leads to the obtainment of superior results in the under-sampling acquisition scheme.

In [21] a dynamic MRI technique in single-plane acquisition through utterance repetition, by reconstructing pseudo-time-varying images of the vocal tract along with a simultaneously recorded audio signal is described. According to the authors, dynamic MRI is a potentially useful tool, allowing one to overcome the main disadvantages of this imaging technique, such as the limited image acquisition time.

In a study described in [22], by using a flashing light synchronized with an external trigger (cardiac gate), the subjects synchronized the utterance with MRI acquisition to assess the feasibility of the imaging of the velopharynx of adults during the repetitive speech production process. Four consonant-vowel syllables were evaluated at a single-slice location ranging from 22 to 29 seconds. Next, in [23] the tongue's shape and dynamics from MR images using manual synchronization of speech and audio data are analyzed.

Other studies have been carried out by using offline reconstruction, from audio simultaneously recorded during the imaging acquisition process. The key components of a speech sequence are identified by the associated position in Fourier space. Due to the high intensity of noise present during MRI acquisition, simultaneous speech audio recording continues to be a challenging problem, as reported in [3]; additionally, the sound input of the intercom does not provide the quality required for an accurate segmentation of the speech signal.

As verified by [24], the noise and speech waveform recorded at the end of the acquisition of an MRI sequence, reveal the high intensity of MR noise produced during the acquisition process, which can change the resultant speech waveform. According to the authors, the relatively loud and disturbing noise produced during image acquisition is regarded as an important drawback of MRI in speech research.

Recently, alternative systems have been presented in order to obtain high quality speech recordings for the subsequent analysis and modelling of the acoustic-articulatory relation from MR images, using an optical microphone [25] or a noise cancellation method [26].

Based on a stroboscopic technique (temporal aliasing applied to a visual phenomenon), in [27] a rapid MRI capable of acquiring a sequence of images constructed from multiple productions of the same utterance (170 repetitions for each sound) through MR scanner noise based synchronization is introduced. In this approach, some constrains may be identified regarding the efforts and mistakes made by a subject due to the numerous repetitions required, in addition to the MR noise non-periodicity.

Initial results on impaired articulation following tongue surgery and reconstruction were presented in [9]; revealing the immediate clinical relevance of the study in terms of compensatory articulatory strategies employed by the patients with articulatory impairments. As demonstrated in former studies, the combined use of acoustic recordings during an MR rapid image sequence allowed for the collecting of useful data in the evaluation of tongue and velum's movements.

In [28] a novel non-invasive imaging technique called tagged Cine-MRI is presented capable of inferring muscle activity within the tongue from tissue deformation. "Tagging" is a method which aids the tracking of the motion of objects along MR image series, and it is particularly useful in cardiac imaging for the quantitative assessment of myocardial's strain or deformation. In the MR images obtained, the "tag lines or grid" appear as dark regions creating a pattern of distortion, which is modified as the object moves.

The inexistence of a specific-protocol or a specific-trigger is the main aspect to be overcomed by dynamic studies. In addition, the quality of the resulting images is still strong conditioned by the necessary compromise between temporal resolution, image acquisition speed and signal-noise ratio.

Several approaches had been developed concerning enhanced MR image sequences and efficient trigger systems. However, the only trigger system currently allowed in clinical practice and conceived for use under the high magnetic fields is the one used by electrocardiogram (ECG) monitoring.

Recent MRI systems allow for rapid image acquisitions with increasing temporal resolution (over 20 frames/sec) for various clinical applications. Most of those real-time applications [5, 12, 17] are based on ultra-fast imaging sequences with different acquisition speed strategies; in particular, the use of K-space (temporary image space for storing data) spiral acquisition or partial Fourier acquisition. Rapid MRI can be achieved by using multiple echoes to increase the completion of the K-space, or through the preparation of the magnetization process prior to the sequence or alternatively, through the acquisition of only half of the K-space.

However, the technology which has been previously described causes the blurring of the MR signal; moreover, [29] has demonstrated that it seriously compromises the quality of the images acquired, thus making it difficult for it to be applied in vocal tract imaging due to the actual anatomy. which is to be visualized (air-tissue boundary) in addition to the complexity associated with the extraction process of the organs' contours.

2. Methods

According to the regulated safety procedures for MRI, subjects were previously informed about the undergoing imaging exam and subsequently instructed about the procedures to be adopted, whereupon they signed a consent form.

Furthermore, the subjects were trained so as to ensure the correct articulation of the intended Portuguese speech sounds and to achieve an adequate speech/acquisition synchronization.

2.1. Equipment and subjects

The images were acquired with the aid of a Siemens Magneton Symphony 1.5T system. The subjects lay in supine position and a head array coil was used for MR signal receiving. After placing the headphones on the subjects, their heads were immobilized with the adjusting coil pads (so as to prevent any involuntary movements). Finally, the MR-electrodes for ECG monitoring were placed on the subjects.

An attempt was made to simultaneously record the subjects' speech during the image acquisition by using the imaging system's intercom. However, this attempt did not guarantee the quality required for the study purposes and thus it was rejected.

The four subjects (two males and two females) who undertook the experiment are young volunteers, without any kind of speech disorders. In order to reduce the required training time for the experiment, the subjects who had therapy skills were chosen.

2.2. Corpus of speech

For the dynamic study, the four subjects produced several repetitions of three consonant-vowel sequences of Portuguese syllables (/pa/, /ma/, /tu/) during the image acquisition process. These syllables were chosen for the following reasons:

1) Syllable /pa/ and the syllable /ma/ - For the purpose of carrying out a comparative study between oral and nasal consonants;

2) Syllable /tu/ - Due to the fact that the articulation of these two sounds produces different modifications in the configuration of the vocal tract leading to more imaging perceivable results.

2.3. Protocol for MR dynamic study

A dynamic study was performed using an image pre-synchronization technique based on the same principle adopted in MRI cardiac studies, with the modification of a FLASH 2D MR sequence using the patient's heartbeat as a trigger, and considering the following acquisition parameters: repetition time (TR) = 60 ms, Echo time (TE) = 4.4 ms and field-of-view equal to 300 mm.

To trigger the speech production, the subjects were instructed to try to synchronize the utterances of the repeated syllables in question with their cardiac rhythm. This was based on their ECG monitorization through a synchronous sound emission conveyed to them by their headphones. The acquisition of each set of images from a single-slice (sagittal) of 6 mm thickness lasted for about 23 seconds and resulted in a reasonable, although variable, number of images for each sequence, as can be observed in Figure 2. Additionally, one image was obtained for each cardiac cycle with an increasing drift in synchronism from a starting reference point in the cardiac cycle (the peak of the R wave in Figure 2).

The adopted time-undersampling method, which was necessary due to the important time overhead presented by the actual MR equipment in each single image acquisition, was based on the assumption that the successive repeated patterns of the utterances, one for each cardiac cycle, were reasonably stationary in terms of pattern and therefore, a reasonable accuracy rate was expected. This was considered to be the best method for achieving a reasonable coverage of the entire articulatory gesture, with the equipment in question.

2.4. Image analysis and measurements

The image analysis was performed by using Dicomworks® software (version 1.3.5) and consisted of the measurement of four commonly used speech articulation parameters, Figure 3, in speech synthesis systems:

1. Lips aperture (LA): the distance between the upper and lower lips;
2. Tongue tip constriction location (TTCL): the distance between the tongue tip and the palate;
3. Tongue body constriction location (TBCL): the distance between the tongue body surface and the palate;
4. Velic aperture (VEL): velum opening.

3. Results

The above mentioned approach allowed for the reconstruction of a variable number of images according to the duration of the each patient’s cardiac cyclet. An average of 6 images (frames) in the same slice position for each syllable with a resolution of 0.853 pixels/mm, was obtained.

Figure 4 shows the sequences of the movements of the articulators performed by a female and male subject during the repetition of the syllables /ma/ and /pa/. The articulatory configurations of the different sounds can be observed in Figure 4. Image 5 depicts the consonant /m/ and the consonant /p/, which is confirmed by the lips' closure, may be observed in image 6. Finally, the vowel /a/, which is established by the lips' aperture and tongue in the lower positions, is depicted in images 1 and 2 for both subjects. As can be observed, the number of frames is not the most important issue to be taken into consideration, and, in most of the images no distinctive features are to be found. Due to the fact that an acoustic recording was not possible during the acquisition process, the image analysis could not be compared with the speech signals.

It should be noted that in Figure 5, one of the sequences resulting from the repetition of the syllable /tu/, as well as some of the contours of the soft tissues involved, are only partially displayed. In the case of the vocal tract in question presenting small differences, the successive shapes which are produced are extremely difficult to detect visually. Thus, 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 in Figures 5c and 5 f.

The sequence of images which is partially represented in Figure 5 depicts the evolution of the male's vocal tract shape during the production of the syllable /tu/, beginning with the plosive consonant, characterized by a sequence initiating from lips' closure (a), which is subsequently followed by the approximation of the tip of the tongue to the upper alveolar region (b). Following this, the tongue moves retreats with the opening of the mouth so as to produce the vowel /u/ (d), and then initiates the inverse process for the following repetition (e).

4. Discussion

Tables 1 to 3 present the speech articulatory parameters obtained among subjects in order to assess various articulatory gestures. In order to carry out a comparative study between oral and nasal consonants in addition to subject genders, two measures were considered: LA to assess lips articulation and VEL to assess nasality during the utterance of the consonants /m/ and /p/. The vowel /a/ was selected in both contexts seeing that the tongue central position and thus does not significantly alter consonant articulation.

In Table 1, one can observe that for both measured articulatory parameters, the standard deviation (SD) values are higher for the female subject than for her male counterpart. As a result of the decrease of LA values the velic aperture increases, thereby demonstrating the nasality component during the consonant uttering, namely in the case of the female subject. The values obtained for the male subject appear to be more inconstant, which can be related to the adopted time-undersampling method for the image acquisition. Speech audio recording during the image acquisition is thus necessary to obtain a more accurate analysis of the obtained results.

For the syllable /pa/ the articulatory parameters presented in Table 2, demonstrate the opposite reality of the previously described phenomena and, as such, the variance is lower between subject genders. During consonant utterances (lower LA values) the velum opening is inferior. However, the results obtained for both subjects are more variable.

In the case of the syllable /tu/, the articulatory gesture is more visible because of the differences in terms of the movement of the articulators involved. Hence, four articulatory parameters were measured. In Table 3, VEL is the most varied parameter associated with the length of the subject’s breath. TTCL occurs in frames 1, 2, 8, 9 and 10. In comparison with the LA and TBCL parameters, the VEL parameter presents a higher SD, thus confirming the active role of the tongue tip during consonant production. The LA parameter has two major positions: in the first five frames the lips were more closed, a greater opening was verified in the following frames. Additionally, the TBCL values are lower when the TTCL values are higher.

When considering the differences among the subjects, this dynamic study demonstrates the existence of significant variability in sound productions between subjects. This is not only due to individual anatomic differences, but also to the peculiarities of each subject’s movement and gesture control that was considered as being extremely individualized as was duly observed.

5. Conclusions and Future Work

The use of MRI provides very useful and precise morphological information as to the positions and shapes of the various articulators involved in speech production, in addition to their temporal dynamics. However, the existence of some restrictions in terms of the speed of imaging acquisition due to the limited temporal resolution of current equipment and protocols has been noted. In fact, improvements and refinements of protocols are necessary in order to obtain further advantages from the current equipments' capabilities.

This paper tries to surmount the previously mentioned limitations by describing a MR dynamic technique, capable of acquiring images of the articulators' movements during speech production, based on synchronization and triggering of the heart. Additionally, the different organs' shapes in some articulatory positions associated with European Portuguese speech sounds have been analysed in MRI sequences, thus resulting in novel perceptions of the dynamic behavior of the articulators in question.

In the future, the authors intend to extend the work presented throughout this paper to other European Portuguese speech sounds, in addition to researching and testing image sequences acquired by 3T (Tesla) MRI equipment with enhanced temporal resolution.

6. Acknowledgments

The images considered in this work were acquired at the Radiology Department of the S. João Hospital, Porto, thanks to the collaboration of Dr Isabel Ramos (Professor at Faculty of Medicine, University of Porto, and Department Director) and the technical staff, who we are most grateful to.

The first author would like to thank and acknowledge the support and contribution of the PhD grant with the following reference SFRH/PROTEC/49517/2009 from IPP – Instituto Politécnico do Porto and FCT – Fundação para a Ciência e a Tecnologia from Portugal.

This work was partially done in the scope of the project "Methodologies to Analyze Organs from Complex Medical Images – Applications to Female Pelvic Cavity", with the reference PTDC/EEA-CRO/103320/2008, financially supported by FCT.

Finally, the authors would like to express their deep gratitude to Dr. Vanessa Esteves from the University of Porto – Portugal for her valuable help with the English version of this paper.

7. References

1. Baer T, Gore JC, Gracco LC, Nye PW. Analysis of Vocal Tract Shape and Dimensions using Magnetic Resonance Imaging: Vowels. Journal of the Acoustical Society of America.1991; 90(2): 799-828.

2. Badin P, Pouchoy L, Bailly G, Raybaudi M, Segebarth C, Lebas JF, Tiede MK, Vatikiotis-Bateson E, Tohkura Y. Un modèle articulatoire tridimensionnel du conduit vocal basé sur des données IRM. Actes des 22èmes Journées d'Etude sur la Parole. 1998; Martigny, Suisse, 283-286.

3. Demolin D, Metens T, Soquet A. Real time MRI and articulatory coordinations in vowels. Proceedings of the 5th Speech Production Seminar. 2000; München, Germany.

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

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

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

7. Badin P & Serrurier A. Three-dimensional Modeling of Speech Organs: Articulatory Data and Models. IEICE Technical Report (Institute of Electronics, Information and Communication Engineers). 2006; 106(177): 29-34.

8. Behrends J, Wismuller A. A Segmentation and Analysis Method for MRI data of the Human Vocal Tract. Proceedings of the Symposium on Human and Machine Perception in Acoustic and Visual Communication. 2001; Tutzing, Germany, 179-189.

9. Mády K, Sader R, Zimmermann A, Hoole P, Beer A, Zeilhofe H, Hannig C. Use of real-time MRI in assessment of consonant articulation before and after tongue surgery and tongue reconstruction. Proceedings of the 4th International Speech Motor Conference. 2001; Nijmegen, Netherlands, 142-145.

10. Kitamura T, Takemoto H, Honda K, Shimada Y, Fujimoto I, Syakudo Y, Masaki S, Kuroda K, Oku-uchi N, Senda M. Difference in vocal tract shape between upright and supine postures: Observations by an open-type MRI scanner. Acoustical Science and Technology. 2005; 26(5): 465-468.

11. Takemoto H, Honda K, Masaki S, Shimada Y, Fujimoto I. Measurement of Temporal Changes in Vocal Tract Area Function during a continuous vowel sequence using a 3D Cine-MRI Technique. Proceedings of the 6th International Seminar on Speech Production. 2003; Sydney, Australia, 284-289.

12. Demolin D, Hassid S, Metens T, Soquet A. Real-time MRI and articulatory coordination in speech. Comptes Rendues Biologies. 2002; 325(4): 547-556.

13. Engwall O. From real-time MRI to 3D tongue movements. Proceedings of the 8th International Conference on Spoken Language Processing (ICSLP). 2004; Jeju Island, Korea, 1109-1112.

14. Crary MA, Kotzur IM, Gauger J, Gorham M, Burton S. Dynamic magnetic resonance imaging in the study of vocal tract configuration. Journal of Voice. 1996; 10(4): 378-388.

15. Di Girolamo M, Corsetti A, Laghi A, Ferone E, Iannicelli E, Rossi M, Pavone P, Passariello R. Assessment with magnetic resonance of laryngeal and oropharyngeal movements during phonation. La Radiologia medica. 1996; 92(1-2): 33-40.

16. Engwall O. A revisit to the Application of MRI to the Analysis of Speech Production - Testing our assumptions. Proceedings of the 6th International Seminar on Speech Production. 2003; Sydney, Australia, 43-48.

17. Mády K, Sader R, Zimmermann A, Hoole P, Beer A, Zeilhofer H, Hannig C. Assessment of Consonant Articulation in Glossectomee Speech by Dynamic MRI. Proceedings of 7th International Conference on Spoken Language Processing (ICSLP). 2002; Denver, USA, 961-964.

18. Faria IH, Pedro ER, Duarte I, Gouveia CA.M. Introdução à Linguística Geral e Portuguesa. Lisboa: Caminho; 1996. (in Portuguese)

19. Nieto-Castanon A, Guenther FH. Constructing Speaker-Specific Articulatory Vocal Tract Models for Testing Speech Motor Control Hypotheses. Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS 99). 1999; San Francisco, USA, 2271-2274.

20. Demolin D, George M, Lecuit V, Metens T, Soquet A, Raymaekers H. Coarticulation and articulatory compensation studied by dynamic MRI. Proceedings of the 5th Eurospeech 97 Conference. 1997; Rhodos, Greece, 31-34.

21. Shadle CH, Mohammad M, Carter JN, Jackson PJB. Multi-planar dynamic magnetic resonance imaging: new tools for speech research. Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS). 1999; S. Francisco, USA, 623-626.

22. Kane AA, Butman JA, Mullick R, Skopec M, Choyke P. A new method for the study of velopharyngeal function using gated magnetic resonance imaging. Plastic and Reconstructive Surgery. 2002; 109(2): 472-481.

23. Avila-García MS, Carter JN, Damper RI. Extracting Tongue Shape Dynamics from Magnetic Resonance Image Sequences. Proceedings of the International Conference on Signal Processing (ICSP 2004). 2004; Istanbul, Turkey, 288-291.

24. Ventura SR, Freitas DR, Tavares JM. Application of MRI and Biomedical Engineering in Speech Production Study. Computer Methods in Biomechanics and Biomedical Engineering. 2009; 12(6): 671-681.

25. NessAiver M, Stone M, Parthasarathy V, Kahana Y, Kots A, Paritsky A. Recording High Quality Speech during tagged Cine MRI studies using a fiber optic microphone. Journal of Magnetic Resonance Imaging. 2006; 23(5): 92-97.

26. Bresch E, Nielsen J, Nayak K, Narayanan S. Synchronized and noise-robust. audio recordings during realtime MRI scans. Journal of the Acoustical Society of America. 2006; 120(4): 1791-1794.

27. Mathiak K, Klose U, Ackermann H, Hertrich I, Kincses W-E, Grodd W. Stroboscopic articulography using fast magnetic resonance imaging, International Journal of Language and Communication Disorders. 2000; 35(3): 419-425.

28. Stone M, Davis E, Douglas A, NessAiver M, Gullapalli R, Levine W, Lundberg A. Modeling Tongue Surface Contours from Cine-MRI images. Journal of Speech, Language, and Hearing Research. 2001; 44(5): 1026-1040.

29. Rua SM, Freitas DR. Morphological Dynamic Imaging of Human Vocal Tract. Proceedings of the Computational Modelling of Objects Represented in Images: Fundamentals, Methods and Applications (CompIMAGE). 2006; Coimbra, Portugal, 381-386.

FIGURE CAPTIONS

Figure 1. MR midsagittal image (slice) indicating the vocal tract's organs during vowel production.

Figure 2. Schematic representation of the ECG monitoring and synchronization technique in the acquisition of dynamic MRI sequences. (It should be noted that in this representation, the number of shift positions were selected in accordance with a generically based viewpoint.)

Figure 3. The speech articulatory parameters which were measured.

Figure 4. Sequences of MR-images for each consonant-vowel sequence of the /ma/ and /pa/ Portuguese syllables obtain from a female and a male speaker.

Figure 5. Contours extracted from midsagittal images obtained in a dynamic study through the under-sampling of the repetition of the syllable /tu/.

TABLE CAPTIONS

Table 1. Comparison of articulatory parameters for the male and female subjects uttering syllable /ma/.

Table 2. Comparison of articulatory parameters for male and female subjects uttering syllable /pa/.

Table 3. Articulatory parameters measurements for the male subject uttering syllable /tu/.

FIGURES

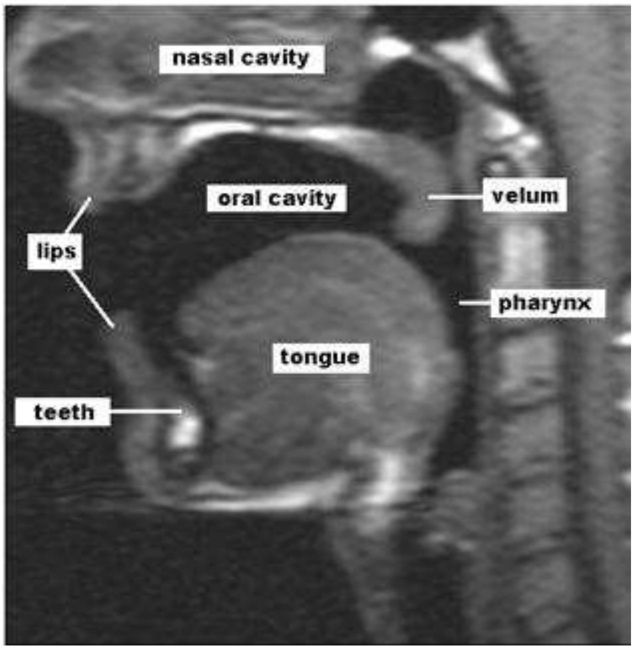

Figure 1

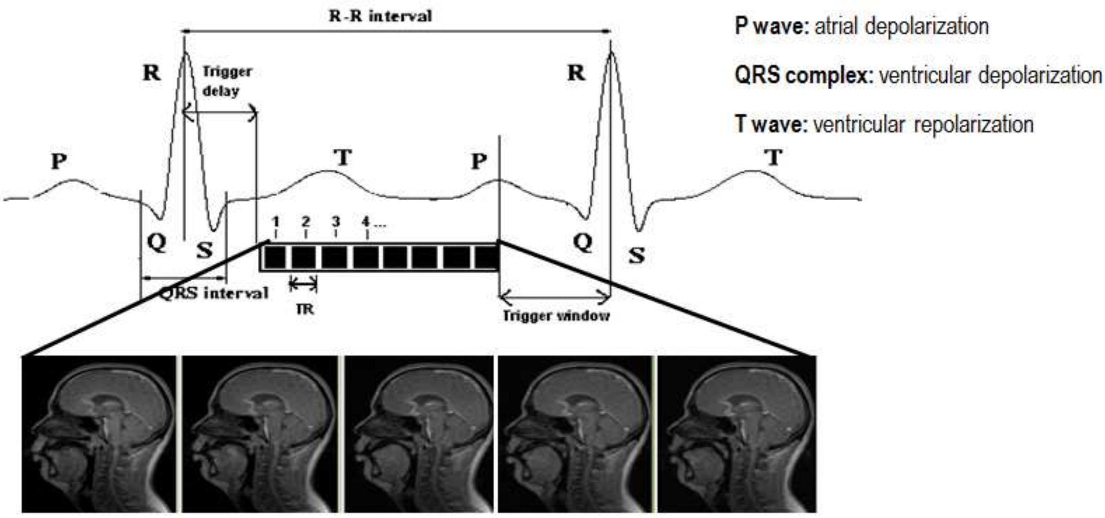

Figure 2

Figure 3

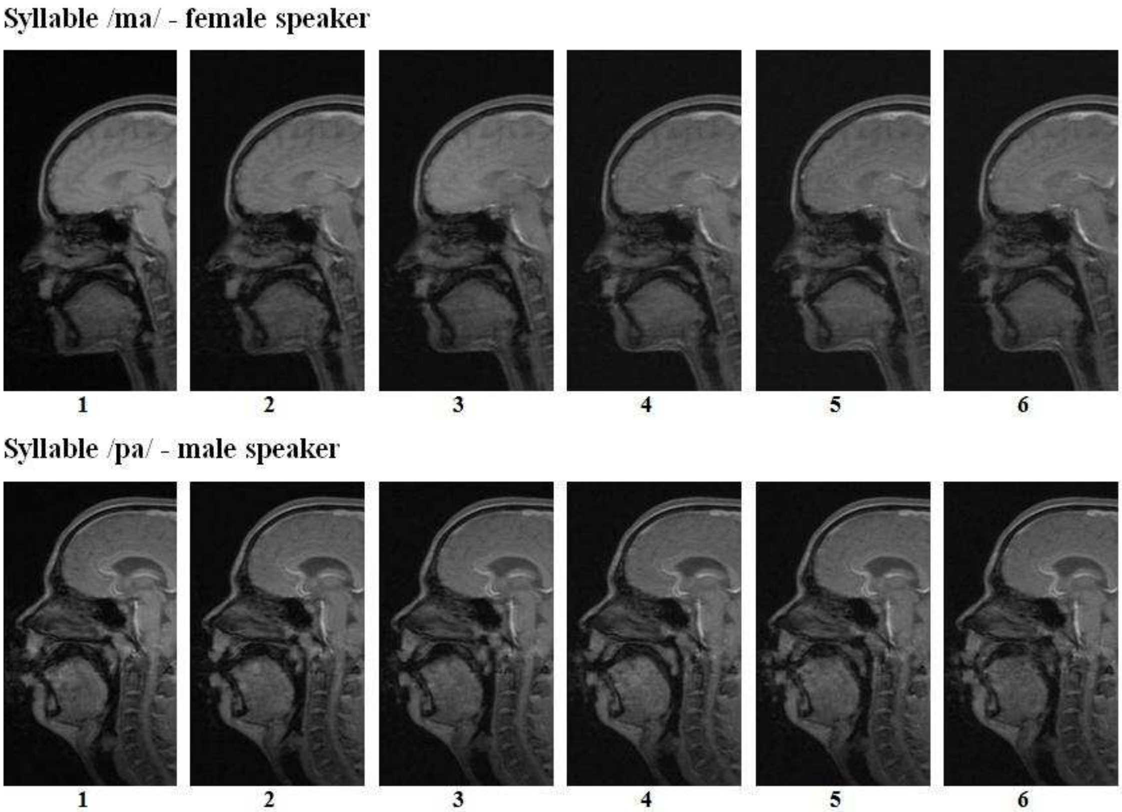

Figure 4

Syllable /tu/ - male speaker

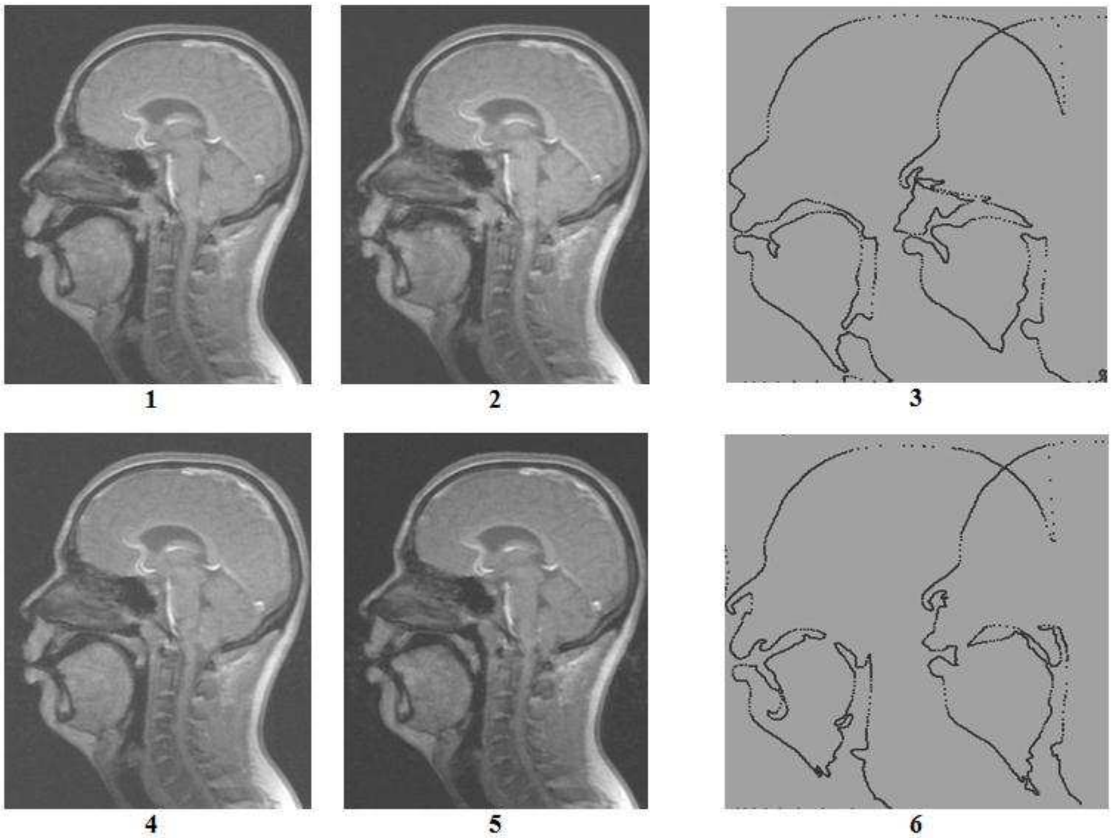

Figure 5

TABLES

Table 1

		Measure [cm] of lip aperture (LA) by frame number								
Subject	Gender	1	2	3	4	5	6	Mean	SD	VAR
1	Male	1.29	1.41	1.76	1.52	1.52	1.29	1.47	0.18	0.03
2	Female	0.82	0.59	0.47	0.00	0.00	0.00	0.31	0.36	0.13

		Measure [cm] of velic aperture (VEL) by frame number								
Subject	Gender	1	2	3	4	5	6	Mean	SD	VAR
1	Male	0.59	0.59	0.70	0.82	0.82	1.05	0.76	0.17	0.03
2	Female	0.59	0.47	0.59	0.82	1.17	1.29	0.82	0.34	0.11

Table 2

Subject	Gender	Measure [cm] of lip aperture (LA) by frame number 1	2	3	4	5	6	Mean	SD	VAR
3	Male	1.29	1.05	1.05	0.94	1.05	1.05	1.07	0.12	0.01
4	Female	1.29	1.17	0.94	0.82	0.59	0.35	0.86	0.35	0.12

Subject	Gender	Measure [cm] of velic aperture (VEL) by frame number 1	2	3	4	5	6	Mean	SD	VAR
3	Male	0.00	0.00	0.00	0.12	0.23	0.35	0.12	0.15	0.02
4	Female	0.00	0.23	0.59	0.47	0.70	0.59	0.43	0.26	0.07

Table 3

Articulatory parameter	Measure [cm] by frame number 1	2	3	4	5	6	7	8	9	10	Mean	SD	VAR
LA	0.35	0.35	0.35	0.35	0.35	0.59	0.59	0.59	0.59	0.59	0.47	0.13	0.02
VEL	0.00	0.00	0.00	0.00	0.12	0.35	0.80	0.70	0.70	0.82	0.34	0.35	0.13
TTCL	0.23	0.35	0.94	0.82	0.70	0.70	0.47	0.35	0.35	0.23	0.51	0.26	0.07
TBCL	0.70	0.59	0.59	0.47	0.47	0.47	0.47	0.47	0.47	0.70	0.54	0.10	0.01