Speech Articulation Assessment using Dynamic Magnetic Resonance Imaging Techniques

S. R. Ventura
School of Allied Health Science – Porto Polytechnic Institute, V.N. Gaia, Portugal

M. J. M. Vasconcelos
Faculty of Engineering, University of Porto, Porto, Portugal

D. R. Freitas
Faculty of Engineering, University of Porto, Porto, Portugal

I. M. Ramos
Radiology Service, St. John Hospital and Faculty of Medicine, University of Porto, Porto, Portugal

João Manuel R. S. Tavares
Faculty of Engineering, University of Porto, Porto, Portugal

ABSTRACT: Magnetic Resonance Imaging (MRI) has been successfully applied on real-time analysis of the articulators during speech production along the whole vocal tract, with good signal-to-noise ratio and without ionizing effects. Because speech dynamic events need a minimal sampling rate, an improvement on the temporal resolution of MRI systems is demanded. Our aim is to describe a dynamic MRI technique to acquire and assess the main articulatory events during the production of some European Portuguese utterances. Hence, novel perceptions for dynamic MRI technique using a 3.0 Tesla System are presented in order to study the shape of the vocal tract during speech production.

KEYWORDS: Image analysis, Medical imaging, Speech production, Dynamic techniques.

1 INTRODUCTION

1.1 *Speech production analysis and challenges*

The speech production mechanism is a complex human motor activity that is able to achieve voice modulation and produce speech based mainly in the articulators' movements. The organs involved, mostly formed of soft tissues, such as the tongue, the lips, the velum and the pharynx, assume extremely important roles during speech production. In fact, these organs together with some bones, i.e. the palate and the jaw, modify the resonance cavities and the shape of the vocal tract in order to produce the sounds.

The human vocal tract's shape (Figure 1) is different among subjects and presents a non-regular contour defined by the air-soft tissues' boundaries. This tube extends from the lips to the glottis, and is formed by four main structures: the oral cavity, the nasal cavity, the velum and the pharynx.

The tongue is the most important articulator, mainly because it is the largest one, and performs a wide range of slow and fast movements during speech production.

Many approaches have been used to track and observe the movements of the articulators, in particular of the tongue, but most of them employ sensors (e.g. electromagnetic articulography) or the direct contact with the tongue and the palate (e.g. electropalatography).

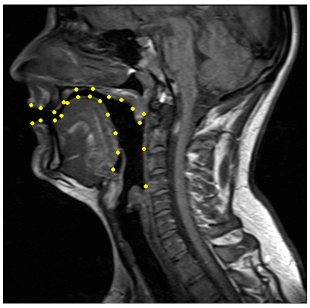

Figure 1. The shape of the vocal tract during the production of [ɛ] vowel in an image acquired by a 3.0 Tesla MR system.

Magnetic Resonance Imaging (MRI) has been successfully applied on real-time analysis of the articulators during speech production, along the whole vocal tract, with good signal-to-noise ratio and without ionizing effects. As such, several MRI techniques provided the calculation of cross-sectional areas and volumes directly from static postures during sustained articulations (Kim et al. 2009) or using multiple repetitions (Stone et al. 2001, Shadle et al. 1999). Because speech dynamic events need a minimal sampling rate and considering the number of articulators involved, two challenges are demanded for the accurate examination, namely:

a) The development of a specific trigger for image acquisition to improve the temporal resolution of MRI systems;
b) Audio-recording in real-time during MRI acquisitions allowing sound acoustic analysis and imaging relationship.

Our aim is to describe a dynamic MRI technique to acquire and assess the main articulatory events during the production of some European Portuguese (EP) sounds.

The remaining of this paper is organized as follows. In the next section, the description of the MRI protocol, the speech corpus and the image analysis and assessment are described. Then, the articulatory measurements obtained for two subjects by using deformable models are presented and discussed. Finally, the conclusions and future outlooks are pointed out in the last section.

1.2 *Dynamic MRI techniques*

A few methods and applications of dynamic MRI have been presented concerning synchronized sampled method (Parthasarathy et al. 2007) or tagging technique (Stone et al. 2001), and achieving images at rates of 7 to 10 frames per second (Demolin et al. 2006, Stone et al. 2001, Mády et al. 2001, Engwall 2004) and even of 18 (Parthasarathy et al. 2007) and 24 frames per second (Narayanan et al. 2004). According to Shadle (1999), dynamic MRI is a potentially useful tool that allows the tracking of the vocal tract organs' movements. In addition, several morphological aspects can be studied as, for example, the motion of the tongue's surface or contour (Stone et al. 2001), the kinematic parameters of the tongue (i.e. velocity, principal strains) (Parthasarathy et al. 2007), the shape of the tongue (Avila-García et al. 2004) and characteristic distances (i.e. articulatory parameters) (Ventura et al. 2011, Echternach et al. 2010). In this area, dynamic MRI can also allow the assessment of articulatory impairments following surgery to structures of the oral cavity, such as for cancer treatment (Mády et al. 2001).

In a previous work, we presented a technique for the dynamic study of the vocal tract with MRI by using the heart's beat signal to synchronize and trigger the imaging acquisition process (Ventura et al. 2011). Our previous dynamic study revealed the existence of significant variability in sound productions among subjects. This variability 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.

Due to the developments that have occurred in MRI, namely by the use of 3.0 Tesla magnetic fields, new applications and image refinements are expected, and consequently significant improvements on the quality of the data acquired with the articulatory events during speech production. Because speech dynamic events demand a minimal sampling rate around 20 Hz (Narayanan et al. 2004), an improvement on the temporal resolution of MRI systems is demanded, for example, by using k-space sampling strategies or more efficient triggering techniques.

1.3 *Vocal tract modeling*

The implementation of statistical methods to analyse data from speech production has proofed to be valuable in several studies. To name a few, (Harshman et al. 1977) used component analysis to identify a set of articulatory features of the tongue. Later, Maeda (1988) applied factor analysis in order to describe the lateral shapes of the vocal tract and (Stone et al. 1997) employed principal component analysis to examine the sagittal tongue's contours from ultrasound images.

The MRI of the vocal tract, associated with the use of statistical deformable models, has made possible the automatic extraction of the vocal tract's shape from the acquired images and the achievement of articulatory measurements that can be useful in the improving of computational speech models (Vasconcelos et al. 2010a, b).

2 METHODS

2.1 *Equipment, subjects and speech corpus*

The image data was acquired using a MAGNETOM Trio 3.0 Tesla MR system and two integrated coils (a 32-channel head coil and a 4-channel neck matrix coil), with the subjects in supine position.

According to the safety procedures for MRI, a questionnaire was performed for screening several contraindications. In addition, the subjects were previously informed and instructed about the study to be performed and the informed consent was obtained. Two young female volunteers, without articulatory disorders, were trained before the MRI exam to ensure the proper production of the intended sounds, and audio recordings were performed before image acquisition and in supine position.

The speech corpus consisted in two sequences of sounds of EP language, in two different articulatory contexts:
i) Vowel-vowel articulation (VV);
ii) Set of consonant-vowel (CV) articulation during a word utterance.

The first articulatory context included the five oral vowels [a ɛ i ɔ u] and the second, the utterance word /pato/ (the English word "duck", IPA phonetic

transcription [patu]). This choice was made considering the sounds familiarity, to any Portuguese subject, during the speech production, and because these sounds are easy to articulate. Furthermore, a coarticulation study was assessed, in the second utterance, stops on consonant-vowel context.
The two audio spectrograms in Figures 2 and 3 illustrate the two sequences of sounds of EP language recorded from the first subject. As can be verified, the vowels sequence utterance has a total duration of 5.05 seconds, and the word /pato/ lasts 2.41 seconds.

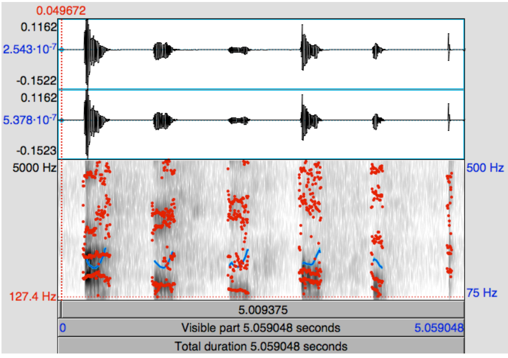

Figure 2. Spectrogram for the first articulatory context in PRAAT software.

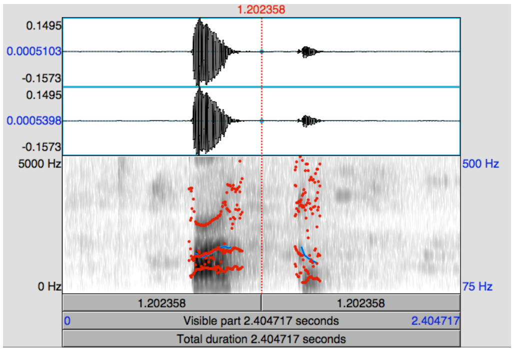

Figure 3. Spectrogram for the second articulatory context in PRAAT software.

2.2 *Dynamic MRI technique*

Concerning the data acquisition, two approaches were used: 1) firstly, rapid pulse sequences and, then, 2) a tagging technique. “Tagging” is a method that aids the tracking of objects’ motion in MRI series. The inserted tags appear as dark regions in the images that move associated to the object under analysis. As such, this technique is particularly valuable in cardiac imaging, as the tissue of the heart’s walls provides few natural features for motion tracking. For speech articulatory assessment, this technique is very difficult to use and more tests must be performed in order to improve temporal resolution and image quality intended for this purpose.
Hence, only rapid pulse sequences combined with parallel imaging were used and tested, through the compromise of the temporal resolution and signal-noise ratio of the MRI system.

Using a Flash Gradient-Echo Sequence, 100 midsagittal WT1 slices were acquired during 48 seconds for each repeated utterance. The MRI protocol parameters used are indicated in Table 1.

Table 1. MRI protocol parameters used in the dynamic study about speech production.

Parameters	*2D Dynamic imaging technique*
TR (msecond)	6.4
TE (msecond)	2.44
Flip Angle	10º
Number of averages	1
Slice thickness (mm)	6
Field of view (mm)	178 x 220
Matrix	156 x 192
Acceleration factor (parallel imaging)	4
Image resolution	0.873 *Pixel* per mm
Pixel spacing	1.146 x 1.146 mm

For the target utterance constituted by the five oral vowels, each sound occurred at least 13 to 37 times per sequence, according to the speed of speech and on the word length. The vowel [a] occurred 27 and 37 times per sequence for each subject. The vowels [ɔ u] were the sounds with lower occurrence rates.
For the second target word, each sound occurred at least 13 to 42 times per sequence, occurring for each subject the vowel [a] about 35 to 42 times and 13 to 16 times for the plosive consonant [t].

2.3 *Deformable models*

In order to perform a more robust analysis of the articulatory behavior during the speech production, statistical Point Distribution Models (PDMs) (Vasconcelos et al. 2008, 2010a) were used to automatically identify, i.e. segment, the vocal tract’s key points in order to compute descriptive measures.
In the building of the PDMs (Cootes et al. 1992), the manual tracing of the key points was carried out by one of the authors with medical imaging knowledge and was realized on images sequentially displayed on the computer screen and later cross-checked by another author. The labeling method was performed according to the anatomic location of the vocal tract articulators (Figure 2).
In the statistical modelling method used, all the training examples are aligned into a standard co-ordinate frame and a Principal Component Analysis is applied to the co-ordinates of the landmark points. This produces the mean position for each landmark, and a description of the main ways in which these points tend to move together (Vasconcelos et al. 2008, 2010a).
The local grey-level behavior of each landmark point can also be considered in the modeling of a shape (Cootes et al. 1993). Thus, statistical infor-

mation is obtained about the mean and covariance of the grey values of the pixels around each landmark point. This information is used to construct the appearance models in Active Appearance Models that can be used to identify the modeled shape in new images.
Active Appearance Models were presented in (Cootes et al. 1998) and allow the building of texture and appearance models. These models are generated by combining a model of shape variation (a geometric model), with a model of the appearance variations in a shape-normalized frame. To identify the modelled shape in new images, the method uses the difference between the current estimate of appearance and the target image to drive an optimization process.

2.4 Image analysis and articulatory assessment

A total of 200 midsagittal MR images were acquired for each subject. Speech articulatory events were described considering seven distances measurements, as depicted in Figure 4. These measures represent the union of the major articulation points for consonants production (because vowels are produced without closure of the vocal tract) and pointed all articulatory (supraglottic) organs, as the lips, the tongue, the palate, the velum, the pharynx.
In order to automatically extract the landmarks of the images, active appearance models were built for each sequence of sounds for each subject (Cootes, 2004). The models were built from the first 20 images of each MRI sequence and later used to automatically label the others 80 images of the sequence.

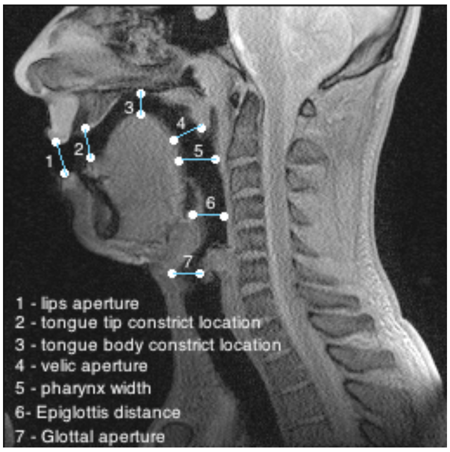

Figure 4. Articulatory measurements addressed on MR images during speech production.

For each subject, MRI data measurements were performed only when an articulatory posture occurred (based on visual assessment of the vocal tract's shape) and, to avoid errors in the analysis and automatic labeling process, rest positions (without speech activity) were excluded.

3 RESULTS

This study presents data concerning the vocal tract's shape during the utterance of sounds of two articulatory sequences and the quantification of seven articulatory parameters.
Considering the large set of images collected, the quantitative results are presented separately for each sound utterance and for each subject.
The analysis was based on the mean values and standard deviations of the seven distances extracted among the subjects, from each image sets that best represent each sound.
The results of the first articulatory context for each subject are presented in Tables 2 and 3.

Table 2. Articulatory measurements of the first articulatory context for the female subject 1.

	[a]		[ɛ]		[i]		[ɔ]		[u]	
	cm	SD	cm	SD	cm	SD	cm	SD	cm	SD
D1	1.051±	0.332	0.821±	0.145	1.055±	0.209	0.897±	0.176	0.872±	0.165
D2	1.363±	0.345	1.232±	0.132	1.048±	0.143	1.364±	0.236	1.130±	0.286
D3	1.274±	0.350	0.951±	0.201	0.635±	0.114	1.184±	0.323	0.927±	0.209
D4	1.288±	0.421	1.316±	0.244	1.488±	0.207	1.101±	0.238	0.949±	0.286
D5	1.154±	0.468	1.528±	0.252	2.137±	0.343	1.112±	0.347	1.226±	0.257
D6	0.984±	0.284	1.253±	0.190	1.606±	0.222	1.215±	0.251	1.126±	0.166
D7	0.901±	0.217	1.175±	0.288	1.108±	0.219	1.105±	0.243	1.044±	0.242

Measurements (cm); Standard Deviation (SD)

Table 3. Articulatory measurements of the first articulatory context for the female subject 2.

	[a]		[ɛ]		[i]		[ɔ]		[u]	
	cm	SD	cm	SD	cm	SD	cm	SD	cm	SD
D1	0.940±	0.118	1.168±	0.087	1.218±	0.078	0.717±	0.184	0.585±	0.112
D2	1.084±	0.132	0.876±	0.073	0.837±	0.112	1.195±	0.083	1.165±	0.088
D3	1.399±	0.279	0.619±	0.181	0.479±	0.135	1.460±	0.111	0.910±	0.222
D4	1.016±	0.225	1.138±	0.161	1.310±	0.168	0.677±	0.127	0.662±	0.138
D5	1.146±	0.306	1.847±	0.141	2.187±	0.249	0.799±	0.108	1.104±	0.173
D6	1.058±	0.147	1.257±	0.175	1.318±	0.149	1.044±	0.126	1.176±	0.058
D7	0.823±	0.262	0.682±	0.107	0.576±	0.098	0.656±	0.084	0.787±	0.129

Measurements (cm); Standard Deviation (SD)

Comparing both subjects, higher distances are demonstrated concerning pharynx width (D5) during the utterance of the vowel [i] and in generally similar measurements can be observed. Except for the open-mid vowels [ɛ] and [ɔ], the global vocal tract's shape (based on the distance trajectory) and distances are fairly different among the subjects.
The results indicated in Tables 4 and 5 represent the measurements extracted from the images collected for the second articulatory context, the word [patu].
Comparing the several measurements obtained for each uttered sound, similar values can be observed for both subjects. Rather different measures are more frequent for each plosive consonant, namely the distances D1 to D3 to the consonant [t] and D2, D5 and D6 to the consonant [p].

Table 4. Articulatory measurements of the second articulatory context for the female subject 1.

	[p]		[a]		[t]		[u]	
	cm	SD	cm	SD	cm	SD	cm	SD
D1	0.626±	0.186	1.002±	0.231	0.410±	0.070	0.650±	0.130
D2	1.046±	0.326	1.389±	0.185	0.599±	0.134	1.050±	0.205
D3	0.999±	0.225	1.232±	0.289	0.790±	0.100	0.887±	0.142
D4	0.901±	0.118	0.975±	0.153	1.063±	0.093	0.775±	0.148
D5	0.933±	0.264	0.789±	0.249	1.139±	0.073	1.069±	0.150
D6	1.043±	0.235	0.860±	0.245	1.260±	0.145	1.234±	0.164
D7	1.047±	0.288	0.832±	0.274	1.015±	0.166	0.943±	0.171

Measurements (cm); Standard Deviation (SD)

Table 5. Articulatory measurements of the second articulatory context for the female subject 2.

	[p]		[a]		[t]		[u]	
	cm	SD	cm	SD	cm	SD	cm	SD
D1	0.649±	0.168	0.887±	0.143	0.643±	0.262	0.620±	0.088
D2	0.851±	0.141	1.033±	0.247	0.460±	0.151	0.872±	0.128
D3	1.102±	0.216	1.250±	0.259	1.019±	0.173	0.963±	0.151
D4	0.838±	0.123	0.937±	0.091	0.933±	0.152	0.805±	0.147
D5	1.108±	0.142	0.976±	0.174	1.179±	0.083	1.223±	0.119
D6	1.172±	0.157	0.995±	0.229	1.138±	0.110	1.317±	0.158
D7	0.952±	0.318	0.793±	0.364	0.825±	0.093	1.126±	0.376

Measurements (cm); Standard Deviation (SD)

Considering the higher occurrence of the target sound [a] (in times per sequence) in both articulatory contexts, a cross-parameter comparison was also performed in order to assess the effects of VV and CV articulation context, Figure 5.

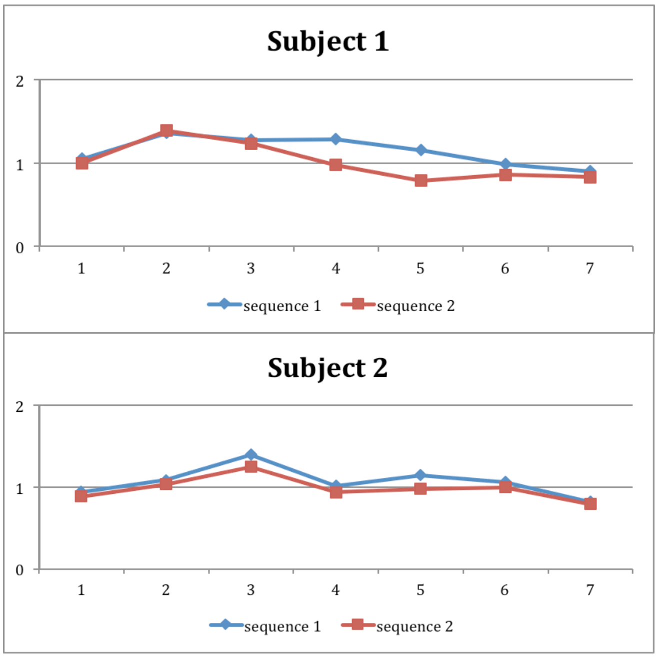

Figure 5. Comparative articulatory measurements performed for each female subject considering the target sound [a] in two different articulatory contexts: VV in sequence 1 and CV in sequence 2.

As can be observed in Figure 5, in general, the values of the distances measured are fairly similar; however, a possible effect of the CV articulatory context in the distances measured, which are clearly lower for both subjects, can be seen.

4 CONCLUSIONS

We have compared the speech articulatory measurements attained in a large set of midsagittal images, acquired during speech production in a very reasonable acquisition time and with enough image resolution for analysis and subsequent quantitative assessment.

Each sound under analysis occurred at least 13 times per sequence during 48 seconds. The major drawback was revealed during the visual assessment of the images for the assemblage of images to each target sound.

By using deformable models in the image segmentation step, we improve the time spend in the process of measurements, as it passes from manual to automatic labelling and, instead of manually labelling 400 images, only 80 were manually annotated.

In line with the acoustic sounds duration, the sound [a] was imaged more times, and the associated measurements show a more linear distribution of all distances (from lips to glottis).

Higher discrepancies of values have been encountered in open-mid vowels, which means the tongue is positioned halfway between a closed vowel (e.g. the vowels [i] and [u]) and an open vowel (e.g. the vowel [a]). This could be related with some errors during the visual assessment of the vocal tract's shape for the assemblage of images to each target sound.

In the word [patu], major differences in the distances extracted were traced for both consonants between the subjects. This can be possible due to the characteristic short time duration of both plosive articulations and result of some measurement errors due to the difficulty of the automatic tracing of the landmark points concerning the lips and alveolar region.

Comparing with our previous works (Ventura et al. 2011 and Vasconcelos et al. 2010a) an improvement on the image acquisition protocol was achieved and more data concerning speech production using MRI was addressed. Hence, a large set of target sounds and articulatory parameters have been analyzed in two different articulatory contexts. Additionally, the statistical deformable models are now used to automatically extract the landmarks of the MR images and thus, to reduce the time employed to extract the trajectory distances of the vocal tract's shape under study.

Articulatory phonetics has long been a discipline with mainly qualitative analysis methods, but in the last decades the advances of MRI allowed the quantification of several articulatory parameters.

The knowledge obtained in this study represents a direct contribution to the improvement of speech synthesis algorithms considering the articulatory parameters measured, and can thereby allow novel perceptions about dynamic behavior of the articulators and co-articulation, namely on European Portuguese speech language.

In addition, these results could encourage researchers and give useful functional and non-invasive information concerning sounds articulation to be used in clinical practice.

In the future, an arrangement for simultaneous recording of speech and MRI of the vocal tract's shape will improve the accuracy of the results, namely for image-acoustic correlation of the target sounds.

5 ACKNOWLEDGMENTS

The images considered in this work were acquired at the Radiology Department of Hospital S. João EPE, in Portugal, and we would like to express our gratitude to the technical staff.

The first author would like to thank the support and contribution of the PhD grant from Escola Superior de Tecnologia da Saúde (ESTSP) and Instituto Politécnico do Porto (IPP), in Portugal.

The second author would like to thank the support of the PhD grant from *Fundação para a Ciência e Tecnologia* (FCT), with reference SFRH/BD/28817/2006.

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 reference PTDC/EEA-CRO/103320/2008, financially supported by FCT.

6 REFERENCES

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* 2: 288-291.

Cootes, T. F., Taylor, C. J., Cooper, D. H. & Graham, J. 1992. Training models of shape from sets of examples in Proceedings of the British Machine Vision Conference, Leeds: 9-18.

Cootes, T.F. & Taylor, C.J. (1993). Active Shape Model Search using Local Grey-Level Models: A Quantitative Evaluation in *British Machine Vision Conference*, Guildford: BMVA Press.

Cootes, T.F. & Edwards, G. (1998). Active Appearance Models. *European Conference on Computer Vision*, Freiburg, Germany.

Cootes, T.F. *Build_aam*. 2004; Available at: http://www.wiau.man.ac.uk/~bim/software/am_tools_doc/download_win.html.

Demolin, D., Sampaio, A. & Metens, T. 2006. An MRI Study of Articulatory Compensation. *In 7th International Seminar on Speech Production*, Brazil: Ubatuba, São Paulo.

Echternach, M., Sundberg, J., Arndt, S., Markl, M., Schumacher, S. & Richter, B. 2010. Vocal Tract in Female Registers - A Dynamic Real-Time MRI Study. *Journal of Voice* 24(2): 133-139.

Engwall, O. 2004. From real-time MRI to 3D tongue movements. *In 8th International Conference on Spoken Language Processing (INTERSPEECH 2004 - ICSLP).* Korea: Jeju Island.

Harshman, R.A., Ladefoged, P. & Golstein, L., 1977. Factor analysis of tongue shapes. *Journal of the Acoustical Society of America* (62): 693-707.

Kim, Y., Narayanan, S.S. & Nayak, K.S. 2009. Accelerated Three-Dimensional Upper Airway MRI Using Compressed Sensing. *Magnetic Resonance in Medicine* 61: 1434-1440.

Mády, K., Sader, R., Zimmermann, A., Hoole, P., Beer, A., Zei-lhofer, H. & Hannig, CH. 2001. Use of real-time MRI in assessment of consonant articulation before and after tongue surgery and tongue reconstruction. *In 4th International Speech Motor Conference*. Netherlands: Nijmegen.

Maeda, S., 1988. Improved articulatory models. *Journal of the Acoustical Society of America* (84 S1): S146-S146.

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.

Parthasarathy, V., Prince, Jl., Stone, M., Murano, Ez. & Nessaiver, M. 2007. Measuring tongue motion from tagged cine-MRI using harmonic phase (HARP) processing. *Journal Acoustical Society of America* 121 (1): 491-504.

Shadle, C.H., Mohammad, M., Carter, J.N. & Jackson, P.J.B. 1999. Multi-planar Dynamic Magnetic Resonance Imaging: New Tools for Speech Research. *In International Congress of Phonetics Sciences (ICPhS99)*. USA: San Francisco.

Stone, M., Cheng, Y. & Lundberg, A., 1997. Using principal component analysis of tongue surface shapes to distinguish among vowels and speakers. *Journal of the Acoustical Society of America* (101-5): 3176-3177.

Stone, M., Davis, E., Douglas, A.S., NessAiver, M., Gullapalli, R., Levine, W.S. & Lundberg, A. 2001. Modeling the motion of the internal tongue from tagged cine-MRI images. *Journal of the Acoustical Society of America* 109 (6): 2974-2982.

Stone, M., Davis, E., Douglas, A.S., & NessAiver, M. 2001. Modeling tongue surface contours from cine-MRI images. *Journal of Speech, Language, Hearing Research* 410: 1-40.

Vasconcelos, M.J.M. & Tavares, J.M.R.S. (2008). Methods to Automatically Built Point Distribution Models for Objects like Hand Palms and Faces Represented in Images. *Computer Modeling in Engineering and Sciences*, 36(3), 213-241.

Vasconcelos, M.J., Ventura, S.R., Freitas, D.R.S. & João Manuel R S Tavares. 2010a. Towards the Automatic Study of the Vocal Tract from Magnetic Resonance Images. *Journal of Voice:* in press.

Vasconcelos, M.J., Ventura, S.R., Freitas, D.R.S. & João Manuel R S Tavares. 2010b. Using Statistical Deformable Models to Reconstruct Vocal Tract Shape from Magnetic Resonance Images. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 224(10): 1153-1163.

Ventura, S.R., Freitas, D.R. & João Manuel R S Tavares. 2011. Toward Dynamic Magnetic Resonance Imaging of the Vocal Tract During Speech Production. *Journal of Voice* 25(4): 511-518.