The Project HERON

António Teixeira[1], *Catarina Oliveira*[2], *Paula Martins*[2], *Inês Domingues*[3] *Augusto Silva*[1]

[1] Department of Electronics, Telecommunications and Informatics, University of Aveiro / IEETA
[2] School of Health, University of Aveiro / IEETA
[3] University of Aveiro / IEETA

ajst@ua.pt, coliveira@ua.pt, pmartins@ua.pt, inesdomingues@gmail.com, augusto.silva@ua.pt

Abstract

In the project HERON, we developed a framework for articulatory speech synthesis for European Portuguese. The system combines several modules, developed or adapted in the project. The Linguistic Processing model uses new syllabification and grapheme-to-phone modules developed in the project. The construction of the gestural scores is performed using an adapted version of TADA (TAsk Dynamic Application), made available for research by Haskins Laboratories. Other important result from HERON was the (first) comprehensive MRI database for European Portuguese and analyses conducted on this data. Two new Electromagnetic Midsagital (EMMA) corpora were also acquired, at GIPSA-Lab in Grenoble, one for EP nasals, and the other, of smaller dimension, relative to EP laterals.

1. Introduction

About 20 ac, HERON, an engineer, decided to give voice to the statue of Memnon, so that it could answer to the caresses of his mother. Since then, or even before, many efforts have been dedicated to give voice to artifacts created by man. Articulatory synthesis produces voice signals using models of physical, anatomical and physiological characteristics of the human voice production system. Even those that use concatenative synthesis, because it is currently the best available method, believe that in the long run that technique is not the answer. In the long term, articulatory synthesis has more potential for high-quality speech synthesis. Since 1995 we have been working in articulatory synthesis of Portuguese, with encouraging results.

The project main lines of research were: to collect new data to support the development of new articulatory models for Portuguese; evolution of the University of Aveiro articulatory synthesizer (SAPWindows); application of new linguistic theories to the development of the necessary processing modules so that the articulatory synthesizer may be the base of a complete text to speech system.

The line of research on the data collection has been divided into two tasks, due to the different data and methodology used. Therefore, project had the following 4 tasks: 1) data collection to support the synthesizer development; 2) development of the synthesizer; 3) development of linguistic modules to enable articulatory based synthesis from text; 4) collection of data and its analyses to support the development of the linguistic models in task 3.

2. MRI Database

Magnetic Resonance Imaging (MRI) was used in order to obtain new anatomic data needed for the development and validation of the articulatory models. This technique has some potential advantages: it provides a good contrast between soft tissues, allows 3D modelling and covers the vocal tract in all of its extension, is non-invasive and considered as safe. Its disadvantages are related to the absence of the teeth in the images, due to their lack of hydrogen protons; the acquisition technique, in which the speaker must e lying down during speech production; the relatively low temporal resolution achieved; the noisy acquisition environment; and the reduced acoustic feedback, due to the use of headphones.

2.1. Main Results

- *A comprehensive MRI database for European Portuguese* [1, 4]. Three different types of acquisitions were performed - 2D static, 3D static, and real-time;
- *Image processing tools for MRI* [3];
- *MRI based studies on European Portuguese (EP) production*: all EP sounds [1, 4], nasals [5] and coarticulatory effects [2].

3. Improvements to the synthesizer

This task included the capability of synthesizing new types of sounds in SAPWindows, starting by adding fricative modeling.

3.1. Main Results

- *Addition of fricatives synthesis* to the SAPWindows articulatory synthesizer [6];
- *Adaptation of the synthesizer to process TADA (TAsk Dynamic Application) output*. TADA is a software implementation of the articulatory phonology approach, developed at Haskins Laboratories. Specifically, TADA is a software implementation of the Task Dynamic model of inter-articulator speech coordination, incorporating also a coupled-oscillator model of inter-gestural planning, and a gestural-coupling model.

4. Linguistic Models

A text-to-speech system based on articulatory parameters must include linguistic processing and conversion from the linguistic discrete variables to the time varying articulatory parameters.

4.1. Main Results

- *A first Articulatory Based Text-to-Speech System for EP*. The system results from the combination of 3 major parts: 1) Linguistic Processing (automatic

syllabification and grapheme-to-phone conversion); 2) the TADA system adapted for EP; 3) synthesizers (the incomplete Matlab CASY implementation, HLsyn and our articulatory synthesizer) [14];

- *Automatic syllabification* [8]. Two different automatic syllabification methods were developed. One uses a finite state transducers (FSTs) approach and is essentially based in the general description of the syllable constituents. The second consists in the implementation of an adapted version of Mateus and d'Andrade syllabification algorithm.
- *Grapheme-to-phone (G2P) conversion* [7]: best results were obtained by a system using a combination of a rule based system (implemented as a finite state transducer) with two machine learning systems (TBL - Transformation Based Learning - and MBL - Memory Based Learning).
- *Preliminary version of Portuguese* TADA [12]. The work included mainly the definition of a gestural dictionary, but also some adjustments to coupling graphs and the use of language-specific dictionary files.

5. Temporal Organization

New corpora were collected and analyses were done by the phonetics researchers for temporal organization of Portuguese articulatory gestures.

5.1. Main Results

- *Acquisition of a new EMMA (electromagnetic midsagittal articulography) database for EP nasals and laterals* [10]. EMMA provides valuable kinematic data relative to different articulators (lips, tongue, jaw, and velum) with good temporal resolution. However, in the majority of available systems, the acquired data are two dimensional and limited to the trajectories of some articulator fleshpoints. The process is invasive and articulation may be affected by the sensors. Recording took place at GIPSA-Lab (Grenoble, France) in October 2007. For each corpus, two subjects were recorded, a male and a female;
- *Automatic annotation of gestures* [11]. Velum, lips and tongue tip gestures were automatically annotated. Velocities for three receivers were automatically calculated in order to determine temporal articulatory landmarks: movement onset, target achievement, target release and release offset. Gestural duration and inter-gestural timing were obtained based on these landmarks;
- *Studies on nasals gestures* [9, 12, 13]. Analyses concentrated on the characterization of gestures duration, velum height and stiffness, and inter-gestural timing. The results confirm the dynamic nature of Portuguese nasal vowels. Results show also clear effects of speech rate on temporal characteristics of EP nasal vowels. Speech rate reduces the duration of velum gestures and increases the stiffness and inter-gestural overlap.

6. Acknowledgements

The project HERON (POSI/PLP/57680/2004) was funded by the Portuguese Research Agency (FCT). Many thanks are due to GIPSA-Lab, Solange Rossato, for help with articulatory design, and Christophe Savariaux, for help with the EMMA recordings. We also thank Radiology Department, Coimbra University Hospital (HUC).

7. References

[1] Martins, P., I. Carbone, et al. (2008). “European Portuguese MRI Based Speech Production Studies." Speech Communication, Volume 50 , Issue 11-12.

[2] Martins, P., I. Domingues, et al. (2008). “Coarticulatory Effects on European Portuguese: A First MRI Study”. M. Embarki and C. Dodane, La Coarticulation : Indices, Direction et Représentation, L'Harmattan: Paris.

[3] Carbone, I. (2008). Segmentação do tracto vocal em dados de ressonância magnética. DETI, Universidade de Aveiro. Mestrado em Engenharia Electrónica e Telecomunicações.

[4] Martins, P. (2007). Ressonância Magnética no Estudo da Produção do Português Europeu. DETI/DLC/SACS. Aveiro, Universidade de Aveiro. Mestrado em Ciências da Fala e da Audição.

[5] Martins, P., I. Carbone, et al. (2007). “An MRI study of European Portuguese nasals”. Interspeech 2007, Antuérpia.

[6] António Teixeira, Roberto Martinez, Luís Silva, Luís Jesus, José Carlos Príncipe, and Francisco Vaz (2005). “Simulation of human speech production applied to the study and synthesis of European Portuguese”. EURASIP Journal of Applied Signal Processing, Special Issue on Anthropomorphic Proc. of Audio and Speech.

[7] Teixeira, A., C. Oliveira, et al. (2006). “On the Use of Machine Learning and Syllable Information in European Portuguese Grapheme-Phone Conversion”. PROPOR 2006, VII Encontro para o Processamento Computacional da Língua Portuguesa. R. Vieira, P. Quaresma, M. das Graças Volpe Nunes et al., Itatiaia, RJ, Brasil, Springer.

[8] Oliveira, C., L. de Castro Moutinho, et al. (2005). “On European Portuguese Automatic Syllabification”. InterSpeech 2005, Lisboa.

[9] Oliveira, C. and A. Teixeira (2007). “On Gestures Timing in European Portuguese Nasals”. ICPhS 2007, Saarbrücken.

[10] Oliveira, C. and A. Teixeira (2007). Nova Base de Dados EMMA relativa às Nasais e Laterais do Português Europeu. Dep. Electrónica e Telecomunicações / IEETA, Universidade de Aveiro.

[11] Oliveira, C. and A. Teixeira (2006). Base de Dados EMMA com Anotação Automática de Gestos. Dep. Electrónica e Telecomunicações / IEETA, Universidade de Aveiro.

[12] Oliveira, C. (2009). Do garfema ao gesto. Contributos linguísticos para um sistema de síntese de base articulatória. Universidade de Aveiro. Dissertação de Doutoramento.

[13] Oliveira, C., Teixeira, A., Martins, P. (2009). “Speech Rate Effects on European Portuguese Nasal Vowels”. Interspeech 2009, Brighton, UK.

[14] Teixeira, A., Oliveira, C., Barbosa, P. (2008). “European Portuguese Articulatory Based Text-to-Speech: First Results”. António Teixeira, Vera Lúcia Strube de Lima, Luís Caldas de Oliveira and Paulo Quaresma (Eds), Computational Processing of the Portuguese Language, Springer.