

# Hélia, Heloísa and Helena: new HTS systems in European Portuguese, Brazilian Portuguese and Galician

Daniela Braga<sup>1</sup>, Pedro Silva<sup>1</sup>, Manuel Ribeiro<sup>1</sup>, Miguel Sales Dias<sup>1</sup>, Francisco Campillo<sup>2</sup>, Carmen García-Mateo<sup>2</sup>

<sup>1</sup> MSFT, Av. Prof. Dr. Aníbal Cavaco Silva, Ed. Qualidade C1, C2, TagusPark, 2744-010 Porto Salvo, Portugal

<sup>2</sup> Signal & Communications Processing Department, Telecommunication Engineering School, Campus Universitario Vigo, 36310 Vigo, Spain

<sup>1</sup>{i-dbraga, i-pedros, i-manrib, midias}@microsoft.com

<sup>2</sup>{campillo, carmen}@gts.tsc.uvigo.es

**Abstract.** Hélia, Heloísa and Helena are the new Text-to-Speech (TTS) systems for European Portuguese, Brazilian Portuguese and Galician, respectively, using state of the art HTS (HMM-based speech synthesis) technology developed at Microsoft. The three TTS systems are presented together as they share most of their modules, namely the same back-end engine and most of the front-end rules. The main differences lie in the language resources. Another difference is the development stage and development goals: Hélia and Heloísa are in RTM (Release to Marketing) version and were integrated in Microsoft products, while Helena is in Beta version and was developed within a R&D project with University of Vigo. In this session, the systems will be shown using the Windows Speech Application.

**Keywords:** HMM-based Speech Synthesis, Text-to-Speech, European Portuguese, Brazilian Portuguese, Galician

## 1 Introduction

Portuguese is among the top 7 languages of the world by number of native speakers [1] and occupies the third place in the ranking of world's western languages. However, the existing TTS systems for Portuguese are still not as many as one could expect, although the main speech technology companies (Acapela, Loquendo, Nuance, Verbio) have at least one available voice for each flavor of Portuguese. Despite the apparent maturity of many commercial TTS systems in Portuguese, with a very acceptable level of intelligibility, most of those systems fail in some front-end related areas, such as in text normalization, homograph disambiguation, out-of-vocabulary words (like foreign words) and opened domain sentences, word stress marks and liaison phenomena. Besides, most of the commercialized TTS systems use Unit-Selection based technology, which may have some disadvantages when compared with HTS: Unit-Selection requires more memory and storage cost (~20-30% more), mainly because of the large voice font, which turns out to be also more expensive, and involves more tuning effort in the phone boundaries' concatenation.

Microsoft developed Hélia and Heloísa, the European (EP) and Brazilian Portuguese (BP) synthetic voices using HTS, a highly intelligible and computationally light technology in order to be integrated in the main Microsoft Unified Communications solutions. Galician, on the other hand, is a regional language spoken in the northwest of Spain, considered by many as an endangered language due to the historical and socio-political influence of Spanish. But Galician is historically and linguistically closer to Portuguese, being considered by some as a variant of Portuguese. This proximity explains the sharing of most of the front-end modules used in the Portuguese TTS systems as demonstrated in [2]. The Galician TTS system Helena is a joint R&D project developed between Microsoft and the University of Vigo with no commercial purposes. Helena will be soon available for demo in a public website.

## 2 System's Description, Tests and Conclusions

The TTS architecture used in these systems was already described for BP in [3]. Since phonetics is the main linguistic difference between EP, BP and Galician, the major investment was made on language resources (lexicon, plain text and corpora with POS tagging, voice font) and testing corpora. Most of the front-end rules could be used from EP with a few adaptations, namely the Sentence Separator, Word Splitter, Text Normalization, Liaison and Homograph Disambiguation modules. The general system workflow, voice talent selection process, back-end and testing procedure was the same. The testing phase represents nearly 1/3 of the product life cycle [3] and ends with an overall voice quality test involving between 20-40 native speakers. In general, listeners preferred HTS-based systems when compared against unit selection-based systems, because the former are highly intelligible and more accurate, especially when dealing with open domain sentences, despite complaining about the well known issues associated with HTS technology: oversmoothing in some sounds, some flat prosody and background buzz. Although Helena (Galician TTS) is in a beta version (which basically means that the front-end modules are not complete), HTS showed much better results than a unit-selection based system. In this demo session, Hélia, Heloísa and Helena will be available for trials using the Windows Speech Application.

## References

1. Lewis, M. Paul (ed.): *Ethnologue: Languages of the World*, Sixteenth edition. SIL International, Dallas (2009) Online version: <http://www.ethnologue.com/>
2. Braga, D.: *Algoritmos de Processamento da Linguagem Natural para Sistemas de Conversão Texto Fala em Português*. PhD Thesis. University of A Coruña, A Coruña, Spain (2008)
3. Braga, D., Silva, P., Ribeiro, M., Henriques, M. and Dias, M.: HMM-based Brazilian Portuguese TTS. In: Braga et al. (eds.) *Propor 2008 Special Session: Applications of Portuguese Speech and Language Technologies*, pp. 47-50. September 10, 2008, Curia, Portugal (2008)