Abstract

In this paper, a linguistically rule-based grapheme-to-phone (G2P) transcription algorithm is described for European Portuguese (EP). A G2P, together with the stress determination and the syllable division, is an essential tool in the general architecture of a Text-to-Speech (TTS) system. The G2P is part of the text pre-processing module of the TTS system and its purpose is to convert text into a phonetic transcription that is interpreted by the synthesis engine.

A complete set of phonological and phonetic transcription rules regarding the European Portuguese standard variety is presented. This algorithm was implemented under the C++ framework and tested by using online newspaper articles. The obtained experimental results gave rise to 98.80% of accuracy rate. Future developments in order to increase this value are foreseen. Our purpose with this work is to develop a module/tool that can improve synthetic speech naturalness in European Portuguese. Other applications of this system can be expected like language teaching/learning. These results, together with our perspectives of future improvements, have proved the dramatic importance of linguistic knowledge on the development of TTS.

The present paper is organized as follows: in section 1, it is made the state-of-the-art on this subject and the justification of our approach; in section 2, the annotation conventions are described, the G2P algorithm is presented and some details on the implementation are shown; in section 3, results are discussed and in section 4 some conclusions and future work are presented.
**Key Words**

**Palavras-Chave**
Conversão Grafema-fone, Regras Fonológicas, Processamento da Fala, Sistemas de Conversão Texto-fala.