

## Automatic Speech Recognition for Portuguese: A Comparative Study

Pedro Henrique Borghi, João P. Teixeira and Diamantino Rui Freitas

Faculdade de Engenharia da Universidade do Porto (FEUP), Research Centre in Digitalization and Intelligent Robotics(CeDRI), Associate Laboratory for Sustainability and Technology, Instituto Politécnico de Bragança

This paper provides some comparisons of Automatic Speech Recognition (ASR) services for Portuguese that were developed in the scope of the Safe Cities project. ASR technology has enabled bi-directional voice-driven interfaces, and its demand in Portuguese is evident due to the language's global prominence. However, the transcription process has complexities, and a high accuracy depends on the ability of capturing speech variability and language intricacies, while being rigorous in terms of semantics. The study first describes ASR services/models by Google, Microsoft, Amazon, IBM, and Voice Interaction regarding their main features. To compare them, three tests were proposed. Test A uses a small dataset with six audio recordings to evaluate in terms of word hit rate the accuracy of online services, with IBM outperforming others (pt-BR:93.33%). Tests B and C utilize the Mozilla Common Voice database filtered by a keywords set to compare online and offline models for Brazilian and European Portuguese regarding accuracy (Ratcliff-Obershelp algorithm), Word Error Rate, Match Error Rate, Word Information Loss, Character Error Rate and Response-Request Ratio. Test B highlights the higher accuracy of Google Cloud (pt-PT:94.90%) and Azure (pt-BR:98.11%). Test C showcases the potential of Voice Interaction's real-time application despite its lower accuracy (pt-PT:78.81%). The tests were carried out using a framework developed using Python 3.x on a Raspberry Pi 4 model B with a server desktop and the REST APIs from the companies' repositories.

**Keywords:** Automatic Speech Recognition · Portuguese · Language Model · Transcription · Mozilla Common Voice · ASR accuracy