A non-Visual Web-Browsing System using Speech Recognition for Brazilian Portuguese

Pedro Batista, Patrick Silva, Nelson Neto, and Aldebaro Klautau

Universidade Federal do Pará, Signal Processing Laboratory,
Rua Augusto Correa. 1, 660750110 Belém, PA, Brazil
{pedro,patrickalves,nelsonneto,aldebaro}@ufpa.br
http://www.laps.ufpa.br

Abstract. This work discusses the integration of available speech technologies for developing assistive tools in Brazilian Portuguese. As a proof-of-concept, it describes an application for non-visual and on-line Web browsing news on the Windows operating system using speech recognition and synthesis.

Key words: Assistive technologies, speech processing, Brazilian Portuguese resources.

1 Introduction

Despite the recognized importance, the Brazilian software industry typically does not benefit from speech technologies in the development of solutions for people with special needs. This work provides a demonstration of how to put together two important assistive technologies: speech recognition and synthesis. The developed application allows a visually and/or physically impaired person to navigate a specific site using his/her voice and listen to the news.

Web navigation for blind people is also possible with the DOSVOX[1] and Firevox[2] softwares. These systems support *screen readers*, which speak out the content of a web page (or the PC screen) to the user using a text-to-speech (TTS) engine. Typically, the interaction between the user and the screen reader is via the keyboard with the definition of shortcut keys. One contribution of this work is to investigate the use of automatic speech recognition (ASR) to improve the user interface.

There are research efforts in the field of Web accessibility to use speech recognition, such as the HearSay project[3]. A related solution is the Simon system[4], which is a language-independent open-source speech recognition program that can be customized to support any application where speech recognition is required, such as web navigation. Both HearSay and Simon do not provide native support to Brazilian Portuguese (BP).

---

[1] http://intervox.nce.ufrj.br/dosvox/
[2] http://firevox.clcworld.net
[3] http://www.cs.sunysb.edu/ hearsay/
[4] http://www.simon-listens.org/index.php?id=122&L=1

This demonstration describes preliminary work on improving non-visual and on-line Web browsing system for BP. The developed application is based on the Microsoft's Speech Application Programming Interface (SAPI) and the Coruja speech recognition system, which was developed by the authors. So, this system allows the establishment of a dialog via ASR and TTS as described in the sequel.

2   Proposed Web-browsing system

The final goal of the project is to make the Web more accessible for visually and/or physically impaired people. The proposed system is domain-specific (customized to specific sites) and adopts a mixed-initiative dialog-based interaction [1]. In the first step, the dialog generator will analyze the Web page content and try to fill a corresponding template with information to be used by the ASR and TTS modules. Each site has its own template. As this project is open-source, any developer can easily add a novel Web site template. The resulting dialog will give the users a more structured view of the page.

The system coordinates the interaction with the user by quickly identifying relevant information (headlines) within the Web site, e.g. the *Folha de São Paulo* newspaper. After this parsing stage, the user is invited to orally choose (via ASR) one of the headlines links. The system then reads aloud the page contents via TTS. The ASR procedure requires the use of speech grammars, which indicate the supported user commands, e.g. skip, repeat and others. The recognition engine confidence level was set to 0.7.

The system currently uses the Nuance Raquel Brazilian Portuguese Female Voice, which supports the SAPI 5.3 interface. This TTS engine is not free. It should be noted that Nuance also provides the freeware L&H TTS3000 engine for BP. However, this latter engine only supports SAPI 4.0, which is not suitable to this project. For the ASR module we used the Coruja package. Coruja is an open-source speech recognition system for BP based on the Julius speech recognition engine[5]. Besides acoustic and language models, the Coruja package contains an API that was developed in the C++ programming language according to the Common Language Runtime (CRL) specification. CRL enables communication between the languages supported by the .NET platform.

The described demonstration intends to substantially improve the browsing experience of people with special needs. The main characteristic is the adoption of ASR via the Coruja speech recognition system, which shows good performance in command-and-control tasks. All resources developed by the authors are publicly available[6].

References

1. E. Horvitz, "Principles of mixed-initiative user interfaces," *Conference on Human Factors in Computing Systems*, pp. 159–166, 1999.

---

[5] http://julius.sourceforge.jp/en/

[6] http://www.laps.ufpa.br/falabrasil/downloads.php