

Open Web-Based Text-to-Speech Services for the Citizens

Spyros Raptis, Aimilios Chalamandaris, Pirros Tsiakoulis, and Sotiris Karabetsos

Institute for Language and Speech Processing – ATHENA Research Centre, Greece
{spy,achalam,ptsiak,sotoskar}@ilsp.gr

Abstract. A system is presented that offers a set of complementary services based on text-to-speech technology. The services and the underlying system that supports them are described. These services include: (a) a service for automatic document-to-speech conversion via e-mail, (b) an open library of audio books, and (c) a dynamic audio news service. The system seeks to maximize the availability and the social impact of text-to-speech technology, making its benefits widely available to the public through open services that address important daily needs of persons with visual impairments and reading difficulties.

Keywords: Speech synthesis, accessibility, open services, audio books, voice-enabled websites.

1 Introduction

Accessibility barriers to information and content, particularly to the ones that are in written form, place significant limitations to the independent living and social inclusion for people with disabilities. As text is well suited for massive and rapid communication, it has become the prevalent means for conveying information, not only in the personal and professional life but also in the relation of citizens to the public administration. As a result, citizens who are disadvantaged in decoding or efficiently handling textual information are in an extremely unfavorable position as they must rely on others to gain access to everyday information that concerns him.

Relevant studies in Greece and other countries have shown that there is a digital divide between the general population and people with disabilities. The limited access to ICT infrastructure, the relatively limited number of websites that meet the W3C accessibility standards and the lack of appropriate training and awareness are three key factors responsible for this digital divide [1]. Some of the measures often proposed for the digital integration of people with disabilities include the development of accessible websites and content relating to news, entertainment, education etc. as well as the participation of the State in providing specialized support software and access equipment.

The work presented here seeks to make a contribution in the above lines, focusing on persons with visual impairments or reading difficulties. They constitute a broad group of persons who are facing difficulties in handling text in written form. Speech is the main alternative to text, and is now considered to be an indispensable component of natural Human Computer Interaction. Its use is rapidly expanding from the

desktop environment to the web and to online content [2-3]. Speech synthesis technology can automatically convert text into speech, thus making text content accessible to reading disabled persons.

The main focus of this work is on making openly available: (a) assistive technologies, based on text-to-speech technology, for serving everyday needs of people with reading disabilities; and (b) content of various types (informational content, news-feeds, educational content etc.) and in various accessible forms.

The next section provides an overview of an online system that incorporates a set of digital services and digital content, and is tailored to the needs of visually impaired persons. This work is carried out under the project "Set of Open, Digital, Speech Services for the Citizens" (MIS: 303620), co-financed by Greece and the European Union.

2 System Overview

In this section, a set of complementary services based on text-to-speech technology and the underlying system that supports them are described.

2.1 The System Core

The core of the system is built on cutting edge *speech synthesis technology* that achieves high intelligibility and near-natural quality. It is based on the speech synthesis engine developed at ILSP/"Athena" and has been ranked at top positions in a recent specialized international contest [4]. The system currently supports Greek and English but due to its open architecture, it can be easily extended to accommodate for any other language. The synthesis engine is supported by extensive abbreviations and pronunciation lexica allowing it to efficiently handle out-of-vocabulary words and other special cases. The engine offers the ability to dynamically modify speaking rate and pitch level.

A *user management module* maintains information about users and their privileges, allowing for different access levels (anonymous, registered and accredited users as well as super users and administrators). Registered users are given personalization options as discussed below. Further to that, accredited users can upload content.

A *content management system* is employed for all the content handled by the system, especially for the digital documents managed by the Library service (Service 2 below).

The overall system architecture incorporates *load balancing* mechanisms targeted to the most computationally intensive part of the processing, namely speech synthesis, thus ensuring a more responsive overall performance.

2.2 The Digital Services

Service 1: Automatic conversion of documents to speech through email

Through this service, anyone will be able to send an email to predefined email addresses, attaching any document in txt, doc or pdf format. The service automatically

converts the attached document to high-quality synthetic speech and returns a reply message with a link to the corresponding audio file. The service offers a degree of personalization, allowing the user to configure and store his/her preferences regarding the characteristics of the synthetic voice to be used in the conversion.

Service 2: Online library of audio content

The Library will provide the basic infrastructure for the storage, management, retrieval and distribution of electronic documents in audio format. It uses a typical metadata scheme for the categorization and organization of content. The Library has built-in mechanisms for the automatic conversion of electronic books and documents in audio form. This way, e-books, magazines, manuals or any document uploaded will be automatically converted and be readily available in the form of an audio file or digital talking book that follows the DAISY standard (Digital Accessible Information SYStem - ANSI / NISO Z39.86). Adopting the DAISY open standard ensures interoperability with content produced or consumed by other technologies and devices.

The Library will offer an initial collection of audiobooks. This will include documents collected specifically for the purposes of the project and will be either free and open content or content for which specific permissions have been obtained. This will cover educational, literary, or other useful books, guides, monographs, conference proceedings, etc. Additionally, the Library will offer acoustic versions of selected public documents and documents from public organizations.

The Library will be open so as to accommodate for an increasing volume of content. Accredited users (public entities, educational institutions, associations of persons with disabilities, etc.) will be given the option to continuously upload new content. This way, the Library will bring significant added value to a range of public or charity institutions providing them with the means to render any document available in alternative, more accessible forms at practically no cost.

The service offers personalization options, through which a user will be able to set a profile of his/her preferences and interests. Based on this profile, users will receive notifications via email for all new content that matches their profile.

Service 3: Audio news alerts

On a daily basis, this service will retrieve news and informational content from different sources and news portals. The content will be presented in the portal in audio form through appropriate novel functionalities which will help users stay informed about current developments in the categories of their interest. Such functionalities include: the option to select and listen to specific news articles, the option to activate live audio news alerts through which the user will be able to listen to headings of news articles in real time as they are published on the portal etc. The news content will be obtained under permission from various sources including both high traffic private news portals as well as public informational websites.

2.3 The Overall Portal

The portal that will host the above services will fully comply with the W3C (WAI/WCAG) guidelines for accessibility at level AA. In addition, in order to further facilitate the accessibility of the target user groups, the portal will be voice-enabled offering functionalities for reading out the content of all webpages in a structured manner through synthetic speech as well as for providing support for voice-enabled navigation [5].

All of the above services will be openly available to all citizens. This way, citizens will gain access to reference content (audio books), dynamic content (news) and the ability to transform their own content into a format that meets their personal daily needs.

3 Progress and Timeplan

The system is currently at the final stage of development. A testing period will proceed, followed by an evaluation phase that will involve users from the targeted groups. The rollout of the final system is expected in autumn 2014. A preliminary limited version of the portal can be found at: www.openspeech.gr (in Greek).

Acknowledgements. The work presented here is carried out under the project "Set of Open, Digital, Speech Services for the Citizens" (MIS: 303620). The project is co-financed by Greece and the European Union, under the Operational Programme "Digital Convergence" and the Regional Operational Programmes.

References

1. Assessment of Digital Divide for Disabled Persons, Immigrants and the Third Age in Greece - Results, Conclusions & Recommendations. Observatory for the Information Society, http://www.observatory.gr/files/meletes/amea_ap_ac.pdf (in Greek)
2. Gilbert, M., Feng, J.: Speech and Language Processing over the Web: Changing the way people communicate and access information. *IEEE Signal Processing Magazine* 25(3), 18–28 (2008)
3. Deng, L., Wang, K., Chou, W.: Speech Technology and Systems in Human-Machine Communication. *IEEE Signal Processing Magazine*, Editors' Note 22(5), 12–14 (2005)
4. Chalamandaris, A., Tsiakoulis, P., Karabetos, S., Raptis, S.: The ILSP Text-to-Speech System for the Blizzard Challenge 2013. In: *Proc. Blizzard Challenge 2013 Workshop*, Barcelona, Spain (2013)
5. Chalamandaris, A., Raptis, S., Tsiakoulis, P., Karabetos, S.: Enhancing Accessibility of Web Content for the Print-Impaired and Blind People. In: Holzinger, A., Miesenberger, K. (eds.) *USAB 2009. LNCS*, vol. 5889, pp. 249–263. Springer, Heidelberg (2009)