Implementation of a polyglot text-to-speech synthesis in two assistive technologies

Paul Fogarassy-Neszly, Aurel Patru
BAUM Engineering
Str. Traian Mosoiu 8, Arad, Romania
pf@baum.ro, ap@baum.ro

Dragoş Daniel Iordache, Costin Pribeanu
National Institute for R & D in Informatics - ICI Bucharest
Bd. Averescu 8-10, Bucucharest, Romania
iordache@ici.ro, pribeanu@ici.ro

ABSTRACT
The text-to-speech (TTS) synthesis is widely used in the area of assistive technologies for visually impaired people as well as for people with dyslexia or related learning disabilities. In the case of multilingual text-to-speech synthesis, both the language identification and voice switching are desirable. This paper presents the implementation of a multilingual text-to-speech in two assistive technologies: the automatic reading machine POET and the portable organizer for blind people Pronto. Both implementations have been tested with visually impaired users. The evaluation results show that the implementation of this software component makes these devices more easy and comfortable to use.

Author Keywords
Accessibility, assistive technologies, usability, technology acceptance, TTS, visually impaired users.

ACM Classification Keywords
D.2.2: Design tools and techniques. H5.2 User interfaces.

INTRODUCTION
A Text-To-Speech (TTS) system is a software component able to produce speech output, whether directly from text input or from a scanned document page previously submitted to an Optical Character Recognition (OCR) system. Many assistive technologies for visually impaired people are using text-to-speech (TTS). TTS synthesis is also used by people with reading disabilities (dyslectic, illiterate, or with learning disability) in order to make the electronic or printed documents accessible. Examples of assistive technologies using TTS synthesis are: screen readers, automatic reading machines, portable computers with voice interface, smartphones, GPS systems, as well as many other gadgets or self-voiced software.

Usually, the synthetic voice corresponding to the native (or preferred) language of the user is currently selected. If the text is written in another language, then the user has to manually select a corresponding voice for that language. In recent years, there is an increasing interest in the applications that are able to process texts written in two or more languages. There are many application areas that need polyglot text-to-speech, such as education for all and multi-cultural contexts [8, 11, 12, 13]. In this case, both a multilingual (polyglot) text-to-speech synthesis and voice switching are needed. This requires to analyze the text in order to detect the language and then to select the voice available for that language. Several approaches for the multilingual TTS exist that differ with regard to the solutions adopted for the text analysis and speech synthesis [1, 3, 9, 10, 11].

In a previous work, a software component for multilingual text-to-speech has been presented that performs both the automatic language identification and voice switching. The component has been developed during the research project iT2V that has been carried on in a consortium of three partners: BAUM Engineering, ETA Automatizari Industriale, and National Institute for Research and Development in Informatics – ICI Bucharest.

The language recognition component plays the role of an intermediate layer, voice independent, between the application and the synthesis process. Language identification is based on statistical analysis and trigrams frequency evaluation for envisaged languages [2]. The development followed four steps: alpha version (proof-of-concept, functional version (beta), commercial version, and implementation in several applications. The development cycle of iT2V is illustrated in Figure 1.

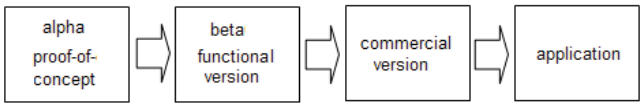

Figure 1. iT2V development cycle.

The goal of the alpha version was to test the language detection algorithms [5]. In the second step, a preliminary functional version (beta01) was developed and tested with four candidate languages [7]. The evaluation revealed a major usability issue: switching the voice in the middle of a sentence. Therefore, is an improved functional version has been developed and tested again. Based on the testing results [6], the commercial version has been developed and then used to implement the software component. The iT2V component plays the role of a special voice (SAPIiT2V) that identifies the language and selects the corresponding synthetic voice.

This paper presents the implementation of the multilingual text-to-speech in two assistive technologies: the automatic reading machine POET and Pronto, the portable electronic organizer for blind and visually impaired users. The implementations have been evaluated for usability and technology acceptance with visually impaired users. The rest of this paper is organized as follows. Next section briefly presents the automatic reading machine POET and the portable organizer for blind people Pronto. Then the evaluation method is presented. In the next section the evaluation results of the implementations in these devices are presented and discussed. The paper ends with conclusion and future research directions.

IMPLEMENTATIONS OF IT2V

Implementation in a reading machine

The first implementation has been done on the automatic reading machine POET (trade mark of BAUM Retec AG).

This type of assistive technology integrates a computer and a scanner (including an Optical Character Recognition software - OCR - with italics adjustment) in one single device. The device is simple and is easy to use. The basic version of POET (Compact 2) has only two buttons (Start and Stop) and two control knobs to adjust the volume and the reading speed. Figure 2 presents the reading device POET Compact 2. The user can read periodicals, books, or magazines. The text could be written on one or several columns.

Figure 2. Automatic reading device POET Compact 2.

The text could be saved and then used for a new lecture or transferred to another device, such as an MP3 player. The model Compact 2+ has three buttons and a menu keypad and has been designed for users that prefer is a model for more advanced users who prefer more features and full control over the functionality. For Braille literate and deaf and blind users a Braille display can be connected to the Poet. As Braille tables are language dependent, prior language detection is necessary before producing Braille output. This is a not foreseen application of language detection.

Implementation in a portable device

The second implementation has been done on the organizer for blind people Pronto (trade mark of BAUM Retec AG).

Pronto is a portable organizer for visually impaired people that runs under the operating system Window CE. The device includes a Braille display and speech output. Pronto supports various applications, like a regular PDA (personal digital assistant). As such is more complex than other assistive technologies. Figure 3 presents the organizer Pronto. The main functions are: taking notes with a text editor, agenda & organizer, playing music, reading books in text or Braille format, Internet access.

The device is easy to handle and easy to use. It provides a keyboard for 8-dot Braille input, an 18-cell Braille display with integrated cursor routing, four function keys for fast access to applications, and a navistick for an easy and comfortable operation.

Figure 3. Organizer for blind people Pronto.

METHOD

Evaluation techniques

The evaluation has two goals: usability and acceptance. First, a usability inspection has been carried on by three experts. Then, each implementation has been tested with visually impaired users for usability and acceptance. After testing the device, participants were asked to answer a questionnaire as regards the actors that are influencing the technology acceptance. The users were also asked to mention the most important positive and negative aspects as regards each device and to weight the relative importance of three factors (ease of use, usefulness, and enjoyment).

Measures

Following measures were collected: number of usability problems by severity degree (major, moderate, and minor), description of unique usability problems, number of positive and negative aspects, the ratings of each item of the questionnaire, and the relative importance of each factor. The items in the questionnaire are presented in Table 1.

Item	Statement
PEU1	It is easy to understand how to use iT2V with this device
PEU2	This device is easier to use with iT2V
PU1	iT2V is a useful facility for me
PU2	The capabilities of iT2V correspond to my requirements
PE1	I prefer to use this device having iT2V
SAT1	I am satisfied with this device having iT2V
INT1	If would like to use this device with iT2V
INT2	I will recommend other people this device having iT2V

Table 1. Questionnaire.

The questions are inspired from the technology acceptance model [4]. Since all the users were visually impaired, the answers have been collected via interview (ratings of each item on a 5-point Likert scale).

Participants and tasks

Seven people participated in user testing (six men and a woman). The mean age of participants is 40.3 years (SD=9.23) with a minimum of 23 and a maximum of 48 years. The user testing took place in Arad, at the Local Branch of Romanian Association of the Blind (Filiala Arad a Asociației Nevăzătorilor din România). Except for one user (university student), all are retired for medical reasons. All participants graduated a high school. The disability degree is severe (first degree - legally blind).

The familiarity with the assistive technologies is variable. However, all are familiar with the lecture on the Internet. The most used information and communication technologies are the computer, mobile phone, tablet, and scanner + OCR. Most used applications are the Internet browsers, Skype, and Facebook. The usual goals are related to information, lecture, entertainment (games), and socialization. The POET device has been tested with three tasks.

The goal of the first task was to read two documents, each of them written in two languages (Romanian and English), without iT2V. This means to stop the lecture and manually change the voices from the device option menu. In the first document the text was written in one column. In the second document, the same text was written in two columns.

The second task was identical, but with the iT2V feature selected. The third task was to read several pages from a magazine with iT2V feature selected. The Pronto device has been tested with a task performed under two conditions. The task was to open and read a document written in two languages (Romanian and English). The condition is to have the iT2V selected.

RESULTS

Reading machine POET

The usability inspection identified several usability problems for the first task. The problems are related to the navigation in the menu in order to manually change the voice. No usability problem has been detected at the second and third tasks. The accuracy of language identification was 100%. The average time saving for a voice switching was 2 minutes. The user testing confirmed the results of the usability inspection: no usability problem has been detected after selecting the iT2V feature.

Item	Factor	Mean	SD
PEU1	Ease of use	4.71	0.49
PEU2	Ease of use	4.43	0.79
PU1	Usefulness	4.29	1.50
PU2	Usefulness	4.29	1.50
PE1	Enjoyment	4.57	0.79
SAT1	Satisfaction	5.00	0.00
INT1	Intention to use	5.00	0.00
INT2	Intention to use	4.71	0.76

Table 2. Descriptives for the POET device.

Most of the users mentioned that the device is accessible, simple, easy to use, useful, and compact. They appreciated the ease of reading and the multilingual feature. As regards the negative aspects, the users mentioned the price (too high), the dimensions (too big) and the lack of the translation.

The users considered that the most important factor for the acceptance of the reading device is the usefulness (45%), followed by the ease of use (29%), and usefulness (26%). The descriptive statistics for the items in the questionnaire is given in Table 2. The mean value of items related to the ease of use is 4.57 (SD = 0.45) and of the items related to usefulness 4.29 (SD = 1.47). An analysis of correlation based on Spearman coefficient shows a marginally significant relationship between the enjoyment and the satisfaction (rho = 0.68, p = 0.091).

The organizer for blind people Pronto

The usability inspection identified few usability problems when performing the task without iT2V. The problems are related to the navigation in the menu in order to manually change the voice. The problems are related to the device itself and somehow inherent, given the rich functionality and the lack of familiarity with the device. The user testing confirmed the results of the usability inspection: no usability problem has been detected after selecting the iT2V feature.

Overall, the usability evaluation results were similar for the two implementations. Most frequently mentioned positive aspects were the portability, the Braille feature, flexibility of use (both voice and Braille), and the multilingual feature. As regards the negative aspects, they mentioned the price and the difficulty to use.

The users considered that the most important factor for the acceptance of the reading device is the ease of use (44%), followed by the usefulness (36%), and enjoyment (20%). The descriptive statistics for the items in the questionnaire is given in Table 2.

Item	Factor	Mean	SD
PEU1	Ease of use	3.71	0.76
PEU2	Ease of use	3.86	0.90
PU1	Usefulness	4.43	1.13
PU2	Usefulness	4.43	0.53
PE1	Enjoyment	4.43	1.13
SAT1	Satisfaction	4.57	0.53
INT1	Intention to use	5.00	0.00
INT2	Intention to use	4.86	0.38

Table 3. Descriptives for the Prontop device

The mean value of items related to the ease of use is 3.79 (SD = 0.39) and of the items related to usefulness 4.43 (SD = 0.53). The analysis of correlation showed a marginally significant relationship between the perceived enjoyment and the satisfaction (rho = 0.72, p = 0.068) and a significant relationship between the perceived enjoyment and the perceived usefulness (rho = 0.84, p = 0.018).

Discussion

The evaluation has been focused on the implementation of the iT2V component and not on the device itself. In this respect, no usability problem has been identified. In both cases, the accuracy of voice switching was excellent.

The component saves time and make the use of these assistive technologies much simpler. The answers to the questionnaires show a difference between the ease of use of the two devices. Pronto has a more rich functionality which makes it more difficult to use. Nevertheless, in both cases the users have been satisfied and expressed the intention to use these technologies.

There are several limitations of this work. First of all, the number of users is small, so the results could be seen as

only exploratory. Second, the users were not familiar with these devices. A reason is the price: six out of seven users said that they can't afford these technologies. Third, the time for evaluation was limited, so only few tasks were assigned. The reasons were the availability of users and the time needed to explain how to use a new device.

CONCLUSION AND FUTURE WORK

In this paper two implementations of the iT2V software component for multilingual text-to-speech and automatic voice switching have been presented.

The evaluation results showed that iT2V is usable, useful and enjoyable. A device with iT2V is easier to use. Since the users are visually impaired, manually changing the voice could be done via an audio menu, which is both difficult to use and time-consuming.

In the next future the language recognition will be implemented in COBRA, the screen reader software. This will be very useful, especially for internet browsing, when beside pages in local language, pages in English should be accessed. Also, in countries with two or more official languages this facility will be useful, as well as for any polyglot user who often are switching between documents in different languages.

Also, beside the language dependent speech synthesis switch, the Braille table will be selected accordingly, taking into account that not only synthetic voices, but also Braille tables are language dependent. This will allow the user to access documents in different languages, both through speech synthesis and Braille display, without any option menu exploration.

ACKNOWLEDGMENTS

This work is partly supported by the IT2V research project (29DPST/2013), financed by UEFISCDI under the PNCDI II Innovation Program.

REFERENCES

1. Bourlard, H., Dines, J., Magimai-Doss, M., Garner, P. N., Imseng, D., Motlicek, P. and Valente, F. Current trends in multilingual speech processing, *Sadhana* 36(5), (2011), 885-915.
2. Cavnar, W. and Trenkle, J. N-gram-based text categorization. *Proceedings Document Analysis and Information Retrieval* (SDAIR-94), (1994).
3. Chen, C. P., Huang, Y. C., Wu, C. H. and Lee K. D. Polyglot speech synthesis based on cross-lingual frame selection using auditory and articulatory features. *IEEE/ACM TASLP* 22 (10), (2014), 1558-1570.
4. Davis, F.D., Bagozzi, R.P. and Warshaw, P.R. Extrinsic and intrinsic motivation to use computers in the workplace. *Journal of Applied Social Psychology*, 22 (14), (1992), 1111-1132.
5. Fogarassy-Neszly, P. and Gherhes, V. Applications for dynamic language identification. *Proceedings of RoCHI 2014* Conference, Popovici D., M. & Iordache D.D. (Eds.), (2014), 51-54.
6. Fogarassy-Neszly, P., Zinveliu, Z. and Pribeanu, C. A software component for polyglot text-to-speech synthesis: user interface and beta testing results. Proceedings of RoCHI 2015 Conference, Dardala, M., Rebedea, T.E. (Eds.), (2015), 145-148.
7. Pribeanu, C. and Fogarassy-Neszly, P. Beta testing of a dynamic language identification software component - preliminary results. *Revista Romana de Interactiune Om-Calculator* 7(3), (2014), 259-272.
8. Ramani, B., Actlin Jeeva, M.P., Vijayalaksmi, P., and Nagarajan, T. Cross-lingual voice conversion-based polyglot speech synthesizer for Indian languages. *Proceedings INTERSPEECH,* (2014), 775-779.
9. Romsdorfer, H. and Pfister, B. Text analysis and language identification for polyglot text-to-speech synthesis. *Speech Communication* 49, (2007), 697-724.
10. Shiga, Y. and Kawai, H. Multilingual speech synthesis system. *Journal of the National Institute of information and Communication Technology* 59(3/4), (2012), 21-28.
11. Traber, C., Huber, K., Nedir, K., Pfister, B., Keller, E. and Zellner, B. From multilingual to polyglot speech synthesis. Proceedings of EUROSPEECH, Budapest, Hungary, (1999), 835–838.
12. Tripathi, M., and Shukla, A. Use of assistive technologies in academic libraries: A survey. Assistive Technology, 26(2), (2014), 105-118.
13. Turunen, M. and Hakulinen, J. Mailman-a multilingual speech-only e-mail client based on an adaptive speech application framework. In *Proceedings of Workshop on Multilingual Speech Communication* - MSC 2000, (2000), 7-12.
14. Udvari-Solner, A. and Thousand, J. S. Creating a responsive curriculum for inclusive schools. Remedial and special education, 17(3), (1996), 182-191.