DEMO & VIDEO

Gary: Combining Speech Synthesis and Eye Tracking to Support Struggling Readers

Gianluca Schiavo
Simonetta Osler
FBK
Trento, Italy
gschiavo@fbk.eu
simonetta.osler@gmail.com

Nadia Mana
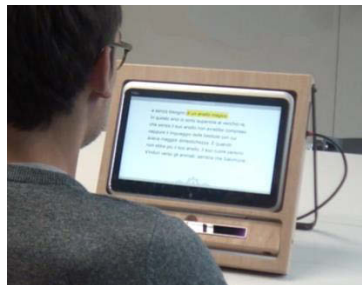**Ornella Mich**
FBK
Trento, Italy
mana@fbk.eu
mich@fbk.eu

Figure 1: GARY is a tablet application for supporting reading that combines speech synthesis and eye tracking.

Abstract

Children with reading difficulties face several obstacles in learning to fluently read written material. Multimedia applications integrating text-to-speech (TTS) synthesisers are valuable tools for supporting reading activities. The paper presents GARY, an application that combines TTS synthesis with eye tracking. GARY is meant to be used on a tablet device coupled with an eye tracker. Making use of the information from reader's eye movement, the system allows users to adapt the speed rate of the synthesised voice to their pace of reading. The paper describes the system, its functioning and future steps in designing a tool for supporting readers' ability in making the connection between the sounds heard and the letters read.

CCS Concepts

• **Human-centered computing~Accessibility systems and tools;** • Human-centered computing~Ubiquitous and mobile computing systems and tools

Author Keywords

Reading Difficulties; Eye Tracking; Speech Synthesis; Technology-Assisted Reading; Accessibility

Supporting reading with multimedia

Technology is a valuable tool for students who struggle to read. Research has demonstrated that specific training programs and multimedia technology support can improve reading performances and content comprehension. As a result, many students with reading difficulties use technology that integrate:

- **text-to-speech synthesizer**, allowing automatic reading;
- **speech recognition applications**, which support the process of typing and spelling;
- **multimodal dictionaries** that provide not only the definition but also the audio pronunciation and the picture of specific words.

Introduction

Learning to read is not a natural process like learning to speak. Reading is a complex process that requires many cognitive skills, such as attention, memory, visual and phonological processing [5]. Learning to read requires mastering and integrating this set of complex skills, and that might not be an easy process for everybody. A large part of pupils in primary education encounter difficulties in reading with fluency and accuracy [8]. While these difficulties might be naturally overcome after completion of the primary education, for a portion of students the obstacle in reading might remain and even characterizes a specific learning disability in reading, i.e. dyslexia [9].

Students with reading difficulties are encouraged to adopt technological tools for supporting them in task involving reading [3]. Text-to-speech (TTS) synthesizers, that allows to automatically read text from digital documents and webpages, are one of the technical tools most commonly used.

Although children are advised to use TTS software for compensating their reading difficulty, TTS applications presents some limitations in their usage, especially for young readers. For instance, the voices used in TTS synthesizers are modelled from adult voices and sometimes not properly suited for young readers [4]. Synthetic voices do not always respect the natural flow and prosody of speech, two characteristics of voice feedback that are particularly supportive in learning to read. Moreover, the speed rate of the synthesis is usually too high for children of primary school age [4].

Traditional TTS applications allow to manually calibrate the voice speed rate, but they offer a very limited set of possible options (usually only three: low, medium and high). Moreover, manually calibrating the voice speed while reading might increase the effort required for the activity, distracting the users from the reading task.

Using eye tracking to track the flow of reading in real time has been studied quite a lot for over a decade. Many eye-tracker applications have been developed for augmenting the reading experience [1] or assisting users with motor disabilities or visual impairments [2]. Nevertheless, the use of gaze tracking for supporting struggling readers has not been extensively investigated, nor dedicated applications have reached the mass market. Moreover, reading text from mobile and tablet devices is becoming a common practice, both for pleasure or for educational purposes. Research has shown that the use of mobile device might improve reading performances when compared with traditional presentation on paper, particularly for a part of struggling readers [6].

GARY - Combining Speech synthesis and Eye Tracking

This paper presents GARY (Gaze and Read it by Yourself), an application for supporting reading with TTS software by combining text-to-speech synthesis with eye tracking technologies (Figure 1). The current prototype is an intermediate step towards an application for tablet devices equipped with an eye-tracker (Figure 2).

Layout personalization

GARY interface can be personalized to meet the preference of the reader. Styles and layout options follow guidelines for presenting written material to readers with reading difficulties.
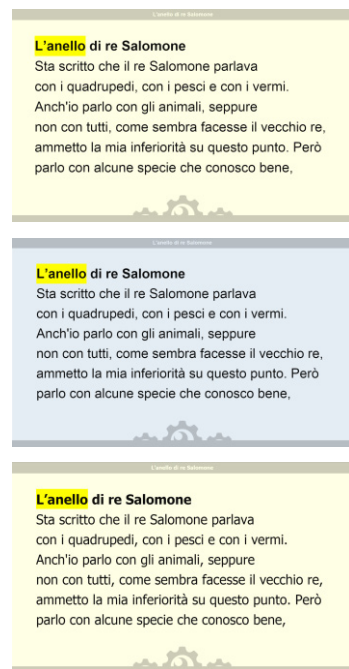
Figure 3: Three layout options: changing background color and font type.
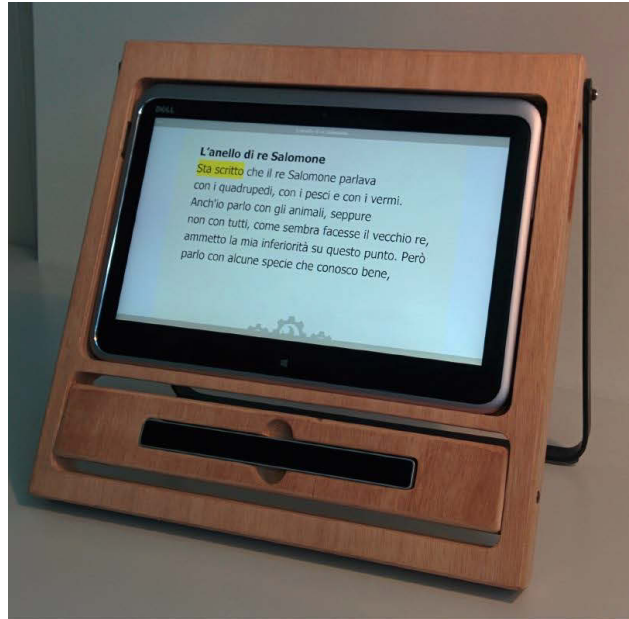

Figure 2: The current prototype (left) is composed of a wood stand that holds a 10" tablet and an eye tracker. In future refinements of the prototype we will aim at integrating the eye tracker on the tablet itself (as seen on the mock-up on the right).

The current application is web-based and meant to be used with an eye tracker (the prototype uses the EyeTribe tracker[1]) for passing the reader's eye coordinates (x and y gaze positions) to the TTS synthesis. The prototype's interface is designed to visualize the text, to play a speech synthesis and to personalize the text layout, following guidelines for presenting written material to readers with reading difficulties (Figure 3). Empirical evidence from psychological research has demonstrated the positive effect that certain manipulations of the perceptual

features of the text may have on speed, accuracy and reading comprehension in struggling readers [7, 10]. Among e-reader and tablet applications, many functions have been proposed for formatting the text in customized ways. These functions include modification of the font, rearrangement of the page layout and manipulations of the dynamics of reading. Specifically, GARY interface allows to change (i) font size, (ii) font family, (iii) background and text colors, and (iv) inter-line and letter space.

GARY integrates a text-to-speech system that automatically turns digital text into speech. The TTS

---

[1] www.theeyetribe.com

system is based on a pre-recorded audio file, containing the reading aloud of the text. The read text is segmented in words (single words or group of words), annotated with the timestamps related to the audio track. As most traditional TTS systems, the text is automatically highlighted during the reading, word by word. The highlighting can help the reader to better follow the voice with attention while focusing the attention on the written word.

The main feature of our prototype is the addition of an eye tracker that permits to gather information from the readers' gaze. In particular, the eye tracker detects and tracks the movements of the reader's eyes and provides the coordinates of the gaze position on the screen, over time. The application maps the gaze coordinates to the corresponding word or group of words and the TTS voice reads the text on which the reader's gaze is directed. In order to deal with inaccuracies of eye tracking measurements, fixations are considered in an area that extends the underlying text size by a factor that could take into account the tracking error. GARY allows users to adapt the speed rate of the synthesised voice to their pace of reading using their gaze and without manually setting any parameter.

Benefits of using GARY

Initial evaluations with young struggling readers and teachers suggest that GARY improve the reading experience compared to traditional TTS software. The benefits are based on two main functions. Firstly, GARY allows the reader to control the pace of the TTS voice by her/his gaze, which guarantees a more natural and effortless reading experience. Secondly, GARY supports the users in "reading with their ears", that is combining listening and reading comprehension. In using

traditional TTS application, the reader can focus solely on the audio feedback, without the need of following the voice on the visual text. In this situation the use of the TTS functionality only supports comprehension from listening and not from reading. Moreover, if the reader does not follow the written text, it might be difficult to locate information on the digital text in subsequent readings. GARY couples the audio feedback with the written text, requiring the reader to explicitly gaze the text to listen the auditory feedback. GARY might thus support the readers' ability to decode and make the necessary connection between the sounds heard and the letters read. Nevertheless, empirical studies involving struggling readers are still needed to evaluate and quantify the benefits of the system in text comprehension and during daily usage.

Acknowledgements

References

1. Ralf Biedert, Georg Buscher, and Andreas Dengel. 2009. The eyeBook – Using Eye Tracking to Enhance the Reading Experience. *Informatik-Spektrum* 33, 3: 272–281.

2. Andrew T. Duchowski. 2002. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers* 34, 4: 455–470.

3. Katharina Galuschka, Elena Ise, Kathrin Krick, and Gerd Schulte-Körne. 2014. Effectiveness of Treatment Approaches for Children and Adolescents with Reading Disabilities: A Meta-

Analysis of Randomized Controlled Trials. *PLoS ONE* 9, 2: e89900.

4. Michelann Parr. 2008. More than words: Text-to-speech technology as a matter of self-efficacy, self-advocacy, and choice. Ph.D Dissertation. McGill University.

5. Robin L. Peterson and Bruce F. Pennington. Developmental dyslexia. *The Lancet* 379, 9830: 1997–2007.

6. Matthew H. Schneps, Jenny M. Thomson, Chen Chen, Gerhard Sonnert, and Marc Pomplun. 2013. E-Readers Are More Effective than Paper for Some with Dyslexia. *PLoS ONE* 8, 9: e75634. http://doi.org/10.1371/journal.pone.0075634

7. Matthew H. Schneps, Jenny M. Thomson, Gerhard Sonnert, Marc Pomplun, Chen Chen, and Amanda Heffner-Wong. 2013. Shorter Lines Facilitate

Reading in Those Who Struggle. *PLoS ONE* 8, 8: e71161.

8. Sally E. Shaywitz, Michael D. Escobar, Bennett A. Shaywitz, Jack M. Fletcher, and Robert Makuch. 1992. Evidence That Dyslexia May Represent the Lower Tail of a Normal Distribution of Reading Ability. *New England Journal of Medicine* 326, 3: 145–150.

9. Sally E. Shaywitz, Robin Morris, and Bennett A Shaywitz. 2008. The education of dyslexic children from childhood to young adulthood. *Annual review of psychology* 59: 451–475.

10. Marco Zorzi, Chiara Barbiero, Andrea Facoetti, et al. 2012. Extra-large letter spacing improves reading in dyslexia. *Proceedings of the National Academy of Sciences* 109, 28: 11455–11459.