The eyePad - Tom Riddle in the 21st Century

Mostafa El Hosseiny
German University in Cairo, Egypt
es.mostafa@gmail.com

Ralf Biedert, Andreas Dengel
German Research Center for Artificial Intelligence
Germany
fistname.lastname@dfki.de

Georg Buscher
Microsoft
One Microsoft Way
Redmond, WA 98052, USA
georg@gbuscher.com

ABSTRACT
We created a multimodal book reader combining eye tracking, handwriting and speech I/O in a novel storytelling concept. We present a number of scenarios integrated in an ad-hoc story to demonstrate new human-text interaction techniques and reading assists, and report on our user study conducted to evaluate the prototype in the real world. Our results show that the new reading assists were invariably reported as being helpful and entertaining.

Author Keywords
Multimodal interfaces, Eye tracking, Handwriting, Speech I/O, Reading, Storytelling

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces—*Input devices and strategies*; J.5 Computer Applications: Arts and Humanities—*Literature*

General Terms
Human factors, Languages, Experimentation

INTRODUCTION
Considering the ongoing miniaturization of eye tracking devices we assume that their large scale integration into tablets and eReaders is a realistic possibility. These devices usually also include touch screens with the capability of handwriting recognition and speech input and output facilities. Given this scenario we want to explore how a user's interaction with such a future reading device can be enhanced in a natural manner, fusing several input modalities to form new reading assists on the one hand, and to provide new authoring capabilities on the other.

Built on top of our idea of the *eyeBook*[2] and various other prototypes[1, 5, 6, 7], it raised our special interest how new ideas of gaze aware interactive reading can be constructed and facilitated. We present a prototype implementing a number of these ideas. It borrows its main ideas from Tom Riddle's Diary first described in the novel *Harry Potter and the Chamber of Secrets* by J. K. Rowling. Tom Riddle's Diary was a blank journal which Tom Riddle transformed into a magical object. The diary allowed a writer to communicate with the memory of a younger Tom Riddle merely through writing on the journal's blank pages.

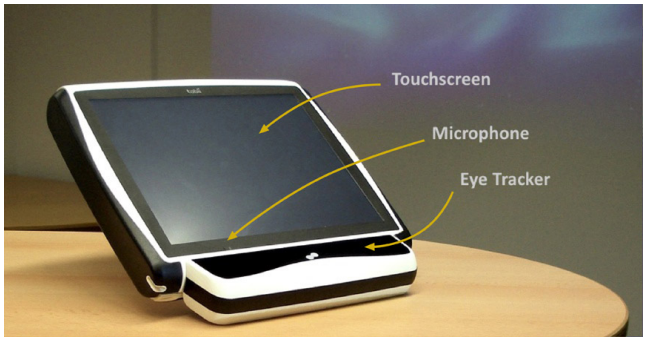

Figure 1. For our interaction research we used a Tobii C12 device. Its main input channels are a touch screen which we use for handwriting recognition, a microphone for speech interaction and an eye tracker to record the user's point of regard.

Numerous work has been done to augment paper books electronically. *Listen Reader*[1] explores sound integration with traditional books. *The mixed reality book*[5] augments book content by adding background music, narrator's voice over, sounds matching pictorial content, animations and augmented surroundings. *The Haunted Book*[7] is an electronic book augmented with animated illustrations of ghost creatures.

However, little work has been done to enhance reading using multimodal interaction techniques. *Novella*[6] is an electronic book reader which combines mouse and speech to navigate and annotate book content. The *eyeBook*[2] is an augmented multimedia book where illustrations, sound effects and background music are adjusted to match the story setting. It also uses gaze control for application interaction (e.g. scrolling).

OVERVIEW AND ARCHITECTURE
We created what we call *eyePad*, a gaze aware, hand writing sensitive, speech responsive prototype that is able to deliver

2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction.
February 13, 2011, Palo Alto, California, USA.

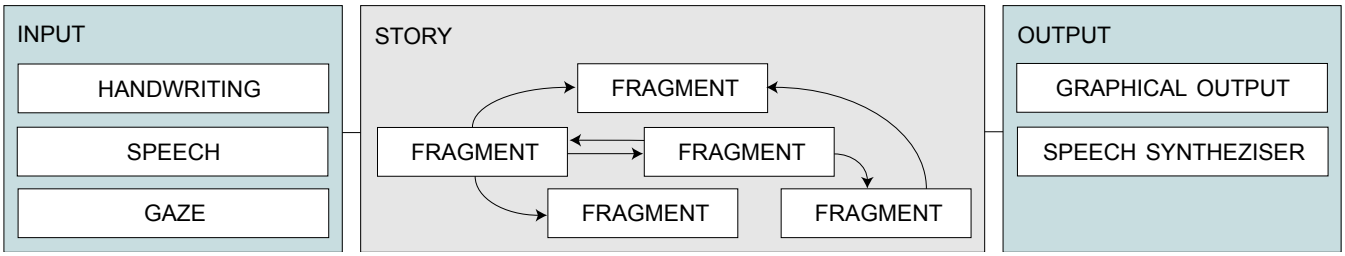

Figure 2. The input evaluators capture the multi-modal input, which is then used by the story module to update the current fragment or to render the next fragment with the help of the rendering/speech facilities.

highly interactive stories[1] to its reader.

The hardware platform is a Tobii C12, see Figure 1, an augmentative and alternative communication (AAC) device intended to be used as an assistive technology for individuals with communication disabilities. The Tobii CEye module is an eye control unit which can be attached to the Tobii C12. The gaze data rate of the Tobii CEye is 30 Hz, and the gaze accuracy is reported to be within 0.8 degrees.

The actual software system is based on a plugin architecture (Figure 2) and fuses the system's main components (compare [4]): input evaluators, the story module, and the rendering and speech output facilities. It is written in Java with the help of Processing[2] and uses parts of the Text 2.0 framework[3].

Input evaluators

The input evaluators monitor the hardware channels and convert the measured *raw* data into high level events which can be facilitated by the story module. In order to convert gaze, handwriting and speech data, we used the following input evaluators:

- Handwriting recognition - Based on an SMO classifier a set of pen coordinates is obtained by observing pen strokes and recording pixel positions for each observation point while it is pressed. A feature vector is then constructed and passed to a previously trained model that is able to perform the classification and emits a list of classified characters. The recognition result is afterwards matched against a set of globally or contextually expected handwriting commands.
- Speech recognition - We built this upon the operating system's inbuilt speech recognition system. Similar to the handwriting recognition the system is primed with a set of expected verbal commands and a callback is executed upon their detection.
- Gaze evaluation - The gaze evaluation module uses the gaze data from the CEye module to keep track of the reader's progress in the text (i.e., which part of the text is currently being read) and if there are regions on the screen that readers can interact with, it keeps track of those that are currently active (i.e., the reader is looking at them). It uses the renderer's layout information and reports input in the form of elements gazed upon.

The Story Module

The story module is the application's heart and drives the story based on the user's input by evaluating story-bundles. A *story bundle* consists of *story fragments*, each containing text, a number of images, expected speech and handwriting commands, and their respective output.

Story fragments have dependencies, that is, only a subset of the story is available to the reader at any moment (of which one story fragment is displayed). Once a fragment has been read or interacted with, the subset changes and the story proceeds according to the user's input and the story's structure.

Due to the nature of story fragments the user's reading behavior is characterized by a high degree of non-linearity, that is, there are several reading paths to choose from and the version of the story that is delivered depends on the specific interactions that took place.

In this respect, the process of reading or interacting with the eyePad is similar to the user's interaction with a *gamebook* or computer game. The player/reader advances in the story by making decisions while the pad dynamically reacts to these choices. The reader may take the decision to perform a certain action, thus triggering a progression in the storyline that is usually irreversible and affects future fragments as well. For example, the pad may ask the reader, "Should X live or die?" causing the reader to respond by writing, "X should die.". The death of X may in turn cause X's mate to seek revenge from the player character or cause the player character to feel remorse for killing X for the rest of the story.

Rendering/Speech facilities

Once the input is evaluated with respect to the current story fragment the output module is employed to display the selected text or synthesize spoken responses. It contains generic renderers for plain text, handwritten text, and images. In addition, special renderers can be implemented as plugins and used for specific types of story fragments.

DEMO SCENARIOS

Using the architecture described above we implemented a story which is explorable by the reader through the afore-

[1]See `http://media.text20.net` for an interaction video with the eyePad.
[2]`http://processing.org`

mentioned kinds of interaction. In addition we integrated a number of special fragments containing novel human-text interaction techniques. The most notable special fragments included:

Interactive map

In the map scenario the reader can follow the protagonist's travel through a wilderness when reading his diary. At the same time, a map (see Figure 3) is updated with suitable icons based on the reader's real time progress in the text. The icons are added transparently and only slightly fade in while his focus is still on the text to minimize distraction. The map served as an interactive visualization and constantly up-to-date remainder of character's location, and the last place of interest is highlighted upon a glance to the left. The reader can also look at any other previously displayed icon and verbally ask 'what is this (again)?' or 'tell me more about this place' to listen to a short abstract. For non-fictional places, like Berlin, the abstract is extracted from Wikipedia, for fictional places, a database has to be populated.

Speaker reminder

Another scenario we looked into was the optimization of dialogs. Many novels contain lengthy conversations between various characters, and frequently the speaker names are omitted due to visual and linguistic elegance, making them hard to follow if one lost track in between. We address this by semantically augmenting dialogs with the actual information about the speaking character. If the reader has no trouble following the text, the book does not trigger any assistance. If the reader encounters, however, any problems during conversations he can gaze on the side of the screen where information about the current speaker is displayed, e.g., an image. In addition, by looking at the image and saying, for example, 'who is she?' the reader can also listen to a brief character profile.

Character reminder

The more general idea of the *speaker reminder* mentioned above was the introduction of a character reminder. The longer a story runs, the more characters are usually introduced and back references can become a source of confusion, especially if they suddenly reappear after some time. Thus we tagged not only the dialog lines with their respective speaker, but we stored a database for every character name and their synonyms. This enables the reader to inquire, for example, 'who was that again?' and likewise receive a brief summary for the character name that the reader momentarily focuses on with his eyes..

EVALUATION

We performed a preliminary analysis of the prototype's capabilities. We designed a user study in which participants interacted with the book and had to perform a number of tasks, the impressions were reported in a questionnaire afterwards. The tasks included:

- responding by handwriting to the book's offers to disclose some parts of its fictitious history.
- reading and interacting with a travel journal, augmented with the map system described above.
- reading a special part of the story that referred to an unknown character which was not introduced before, thus simulating forgetfulness.

Participants

Eight Participants performed all tasks, i.e., four males and four females, with an average age of 21.5 years. They were university graduate and undergraduate students majoring in computer science and engineering. The completion of all four tasks took around 30 minutes (including calibration and training).

Results

Our first question was if the integrated modalities enhanced the reading experience (see Table 1). Although 63% of the participants thought that the pen interaction enhanced their reading experience, one participant mentioned that he would rather tap buttons on the screen than ask/answer questions. 88% of the participants were pleased with the gaze interaction describing it as 'useful' and 75% of the participants thought that the speech interaction enhanced their reading experience.

Figure 3. A disoriented reader looks at the map to find out the last place of interest.

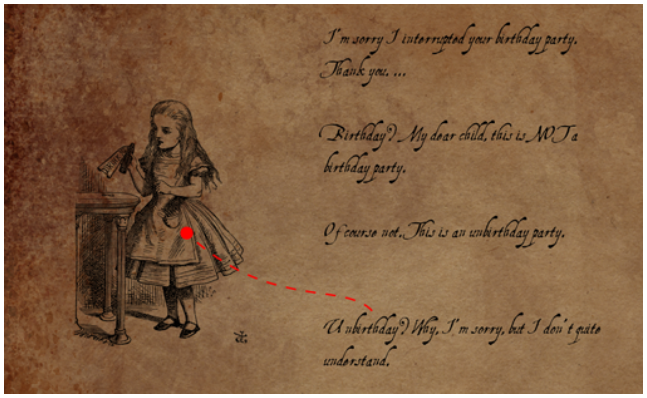

Figure 4. A reader observes an illustration of the unknown speaker after looking to the left.

Table 1. Does integrating pen, speech and gaze input modalities enhance the reading experience?

	Yes	Neutral	No
Gaze interaction	88%	0%	13%
Speech interaction	75%	25%	0%
Pen interaction	63%	25%	13%

Table 2. How do readers find the quality of the eye tracking, the handwriting recognition and the speech recognition/synthesis? Results were reported on a five point Likert scale, 1 equals very bad, 5 equals very good.

	Rating
Handwriting recognition	3.9
Speech recognition	3.9
Speech synthesis	3.5
Eye tracking	3.4

In order to distinguish whether disfavoring ratings were caused by principal interaction flaws or rather caused by a poor implementation of a particular subsystem we also asked the participants for explicit ratings of their perceived performance and accuracy of the individual subsystems. They were asked four questions of the form 'How would you rate the quality of X?' where X is replaced by the subsystem under consideration. The ratings went from 1 (very poor) to 5 (very good), a value of 3 was considered acceptable, see Table 2 for the results.

Additional open feedback could be given as well, in here the most notable points were related to the speech input and output. Some users usually expected their utterances to be recognized, even if they did not speak loudly and in a clear voice. For others the speech synthesis voice was too fast. One participant mentioned that it was difficult to understand, and that it needs a very silent environment. Regarding the eye tracking performance it could be observed that the device's accuracy degraded due to the participants' urge to move during the experiment, resulting in a shifted head position relative to the calibrated position.

We also researched how the readers reacted to the presented scenarios. All participants reported that all of the special assists happened to be helpful. However, although all participants agreed that the interactive map makes the text easier to visualize, one participant mentioned that he did not notice when the icon was placed on the map because he was concentrating too much on the text. Another participant said that it was not clear whether they should look at the map while reading or after reading the text.

The speaker and character reminders were thoroughly received positively, they were reported to be uniquely distinctive to ordinary (e-)books and would really 'add something extra'.

OUTLOOK & CONCLUSION

We presented a multi modal gaze aware, hand writing sensitive, speech responsive prototype on a tablet computer. We implemented a demo story and integrated story fragments containing novel human-text interaction techniques. We also evaluated the prototype's capabilities in a user study that addressed various issues of the interface and the implementation.

Initial results with regards to the overall reading experience and the helpfulness of the proposed reading assists were very satisfactory considering the prototype status. Improving the robustness and quality of the eye tracking, speech recognition/synthesis and handwriting recognition is crucial to the eventual acceptance of the prototype by real users as evident from the results of our user study.

The usage of non-linear story fragments and their implicit control through gaze, or explicit control through handwriting and speech allow for exciting possibilities. Book authors are given an interesting set of plot and interaction means to shape a story according to their own imagination and the reader's progress. We see also plenty of use-cases for our prototype in various domains including e-learning and entertainment.

REFERENCES

1. M. Back, J. Cohen, R. Gold, S. Harrison, and S. Minneman. Listen reader: an electronically augmented paper-based book. In *Proc. SIGCHI conference on Human factors in computing systems*, page 29, 2001.

2. R. Biedert, G. Buscher, and A. Dengel. The eyebook, using eye tracking to enhance the reading experience. *Informatik Spektrum*, 2010.

3. R. Biedert, G. Buscher, S. Schwarz, M. Moeller, A. Dengel, and T. Lottermann. The Text 2.0 Framework. Workshop on Eye Gaze in Intelligent Human Machine Interaction, 2010.

4. A. Gourdol, L. Nigay, D. Salber, J. Coutaz, and L. de Génie Informatique. Two case studies of software architecture for multimodal interactive systems: Voicepaint and a voice-enabled graphical notebook. In *Proc. IFIP TC2/WG2*, volume 7, pages 271–284.

5. R. Grasset, A. Duenser, H. Seichter, and M. Billinghurst. The mixed reality book: a new multimedia reading experience. In *CHI '07 extended abstracts on Human factors in computing systems*, page 1958, 2007.

6. J. Hodas, N. Sundaresan, J. Jackson, B. Duncan, W. Nissen, and J. Battista. NOVeLLA: A multi-modal electronic-book reader with visual and auditory interfaces. *International Journal of Speech Technology*, 4(3):269–284, 2001.

7. C. Scherrer, J. Pilet, P. Fua, and V. Lepetit. The haunted book. In *Proc. 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, volume 0, pages 163–164, 2008.