**Multimodal text Editing**

**Experimental Description:**

The main objective of this experiment is to understand how we can edit voice input as text entry with the help of eye tracking.

Text entry by eye-typing is quite slow in comparison to conventional text entry on keyboard. Speech to Text entry is quite fast. However, its prone to error and noise thereby leading to ambiguous text entry. Just voice based text entry systems do not have any corrective features. However, if voice and eye tracking can be combined, that could lead to an improved and faster experience of text entry with quicker text editing features. Multimodal text entry has always existed on conventional devices (voice + touch/mouse). This is being extended in eye tracking environment to improve the process of text entry. Earlier research focuses on eye tracking as a pointing approach to text insertion by voice commands. Our approach takes into account the limitations of voice and eye to provide a reasonable editing solution in multimodal environment.

We designed an approach where to correct a word that is wrongly inserted, the person is supposed to “look” at the word for a fixed duration (let’s say 1.2seconds). Once the wrong word is selected, there will be an option to edit. At this time, the speech to text option will be paused/closed. In edit mode, one can type over the gaze-keyboard with the help of the eye tracker. The keyboard will have a done button which when activated will replace the wrong word with the right word. While typing, one will be able to see the complete sentence typed. (When the keyboard is activated, there will be option to choose words from text word suggestions as well)

This application will be a static HTML JS CSS page running on a CEF designed browser(GTW).

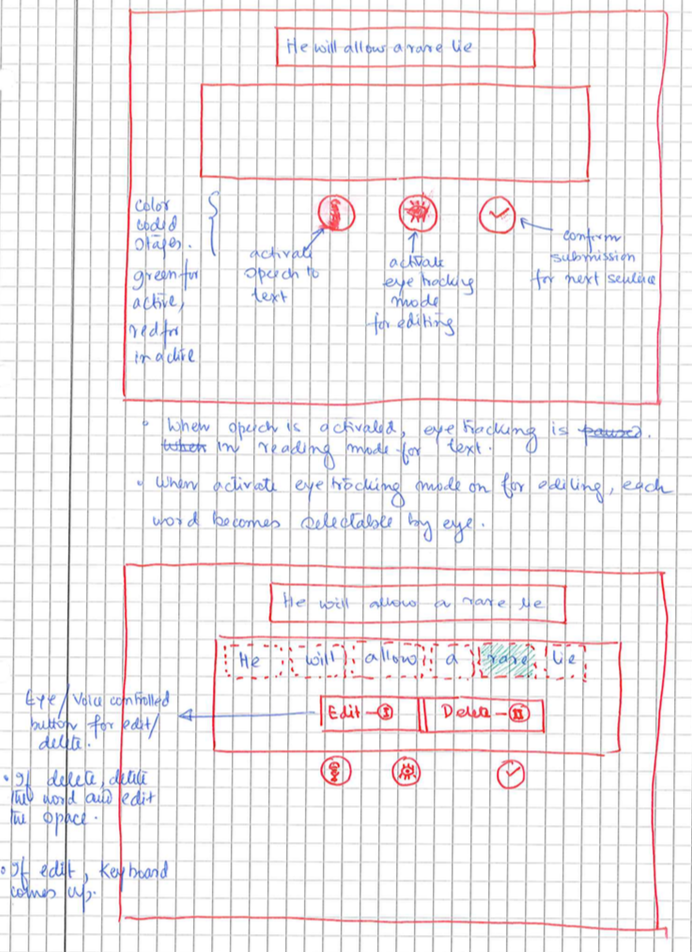
**Measurements**

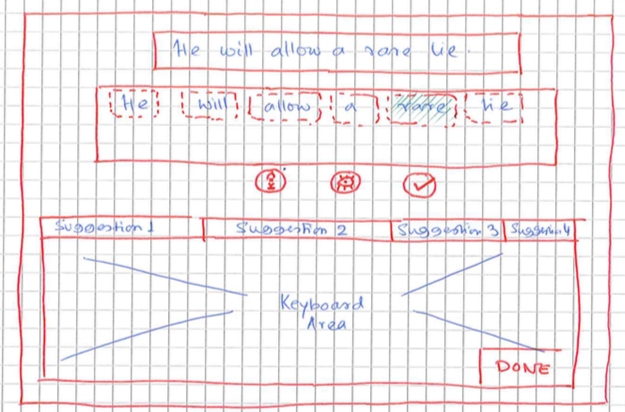
1. Every key interaction that is activated will be logged. For example, if it takes 1.2 seconds to activate the edit menu, we will have to measure how much time it took for editing the complete word.
2. How many words are edited will also be taken into account.
3. The time taken to complete the sentence (with/ without error)

**Documentations**

1. Mouse hover text : <http://jsfiddle.net/shahe_masoyan/z7nkU/1/>
2. Keyboard: <https://mottie.github.io/Keyboard/>
3. Web Speech API: <https://developers.google.com/web/updates/2013/01/Voice-Driven-Web-Apps-Introduction-to-the-Web-Speech-API>
4. Web Speech API example: https://www.sitepoint.com/introducing-web-speech-api/

**Mock Designs**

****

****

**Tentative Plan**

**Phase 1 (August 10)**

HTML page that page a text box and a mic button that has the Speech to Text API implemented. The job is to hover over the button and let it get selected in 1 second. There needs to be a delay as in eye tracking super-fast activation leads to higher error rate. The moment the button gets activated, one can hear a sound indicating that its ready to listen to our speaking.

**Phase 2 (August 20)**

Design of keyboard with similar visual hovering effect. Each key will have a delay of 1 second.

**Phase 3 (September 5)**

Selection of words from text input to go into edit mode.

**Phase 4 (September 15)**

Edit function and replace the entire word with a new word or with delete action.

**Phase 5 (October 6)**

Enter Trial phase with proper measurement of data.

**Phase 6 (October 20**)

Enter experimental evaluation phase. (10 participants)

**Methodology**

**Learning Phase:**

Each participant will have a learning phase of 10 sentences for 2 sessions. They will be instructed on how to edit words in the sentences when they appear with an error from the speech to text application.

**Evaluation Phase**:

Each participant will undergo an experimental session of 10 sentences for 5 sessions.

**Sentence Set**

1. MacKenzie, I. Scott, and R. William Soukoreff. "Phrase sets for evaluating text entry techniques." *CHI'03 extended abstracts on Human factors in computing systems*. ACM, 2003.
2. Garofolo, John S., et al. "TIMIT acoustic-phonetic continuous speech corpus." *Linguistic data consortium* 10.5 (1993): 0.

**Some Related Work**

1. Beelders, Tanya René, and Pieter J. Blignaut. "Using vision and voice to create a multimodal interface for Microsoft Word 2007." *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ACM, 2010.
2. Beelders, Tanya René, and Pieter J. Blignaut. "Measuring the performance of gaze and speech for text input." *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 2012.
3. Hoste, Lode, Bruno Dumas, and Beat Signer. "SpeeG: a multimodal speech-and gesture-based text input solution." *Proceedings of the International working conference on advanced visual interfaces*. ACM, 2012.
4. Beelders, Tanya René, and Pieter J. Blignaut. "Using eye gaze and speech to simulate a pointing device." *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 2012.

**Future Plans**

Master’s Thesis with complete text editing solution.