June 10, 2016

Geoffrey Cooper
KTH Royal Institute of Technology
Stockholm, Sweden
geoffrey.cooper@gmail.com

Benefits of Reading Text in a Multimodal Manner

Abstract
This paper proposes a multimodal reading system and researches the benefits and drawbacks to the system. These benefits and drawbacks were uncovered by studying previous research on the use of multiple modalities in language, redundant modalities, and sensory processing in the brain. Based on the research, the multimodal reading system may have significant benefits, on the condition that the system fits with the brain's way of processing sensory information.

Introduction
Most of the text we read on a daily basis is displayed on a screen, and we read using our visual sense in similar way to reading a book or newspaper. Meanwhile our other senses are idle or might be engaged in other events, such as hearing the ambient sounds in a cafe. When reading on a digital device, it may be possible to engage multiple senses in the act of reading, which could result in a more focused experience and better memory retention.

This would be possible with an eye-tracker that's built into a digital device. If a person's device knows which word they're reading at any given time, it would be able to output sound, colour and haptic feedback that corresponds to that word. Taking sound as an example, the sound of each word could be algorithmically generated, resulting in a unique sound for each word. The sound for the word "apple" would be a unique abstract sound that is heard every time the person read the word "apple". As the person reads through a paragraph of text, they would essentially hear a symphony of sounds that is the auditory translation of that text. In the same way, colour and haptic feedback could also be provided while reading.

Theoretically, after becoming accustomed to the system over time, a person would be able to read text in a multi-modal manner, which could have several benefits. The potential benefits and drawbacks of such a system are studied in-depth in this paper, by conducting a literature review of related topics.

Background
Some prior research exists for the use of multiple modalities in language. The most relevant field to this study might be the teaching of language literacy to children with Dyslexia using multimodal feedback.

One of the most widely used literacy teaching techniques is the Orton-Gillingham Approach, which has been in use since the 1930s. In this approach, the auditory, visual, and kinesthetic input channels are used simultaneously by the students [1]. A very simple example of this is saying a word out loud while simultaneously writing it. The Orton-Gillingham Approach has been tested in many studies, and results

have shown significant improvements over non-multimodal learning techniques. Teachers with dyslexic students have a wide variety of teaching techniques and exercises available, and a large proportion of these utilize multimodal feedback to reinforce learning [2].

Another related field is mnemonics, which are learning techniques that help with information retention. Some common mnemonics utilize multiple modalities in order to remember information. For example, in alphabet songs, sounds are associated with letters in order to improve retention. Daniel Tammet, who holds the record for reciting pi from memory, says that in his mind each positive integer up to 10,000 has its own unique shape, colour, texture, and feel [3]. The mnemonic that Tammet uses is a result of having synesthesia, a neurological phenomena in which the stimulation of one sense is linked with another sense.

From a technical standpoint, the Text 2.0 project led by Ralf Biedert shares some similarities to the multimodal reading system proposed in this paper. The project utilized an eye-tracker to track a person's eyes while reading text on a screen, and displayed additional information to the reader depending on which word they were looking at. For example, staring at a word for roughly one second could make a translation appear above it, or trigger a spoken sound clip of that word [4].

Seeing as related works are few and far between, this paper will investigate existing knowledge about the benefits and drawbacks of combining modalities.

Method

This paper seeks an answer to the question of whether or not a multimodal reading system would be beneficial. Some more specific questions are posed to help guide the literature review:

For a multimodal reading system,

- Would combined multimodal feedback have any benefits over a single mode?
- Would memory/retention be increased?
- Would focus be increased?
- Would there be any negative effects?

An extensive web search was conducted for related topics, such as multimodal interaction, redundant modalities, the use of multiple modalities in language, and sensory processing in the brain.

Results

In the proposed multimodal reading system, when a person reads over each word on a screen, multiple kinds of sensory information would be simultaneously output to the reader. For the word "apple", the reader would see, hear, and feel different signals, but all of the signals have the meaning of "apple". This strategy is known as multimodal redundancy : "two or more modalities are used to communicate the same information" [5].

In the literature review, many positive benefits were found for using multimodal redundancy. First, it makes interactive systems more accessible by allowing people with visual impairments and hearing impairments to choose their least error-prone modality. According to a

researcher at KTH, "Multiple modalities are very important because a deficit in on channel of communication may be overcome by exploitation of another" [6]. This works for people who are situationally impaired as well. One sense might be occupied or distracted, and the person can easily switch modalities as needed. Multimodal redundancy also increases robustness. An example of this is Gabriel Skantze's speech synthesis software, which was used to test listening comprehension. The first group of participants heard only the audio of the words. The second group heard the audio and also saw a virtual face on the screen, whose lips formed the words. The second group who received both visual and auditory feedback were able to guess twice as many words correctly [7]. This finding correlates with the previously discovered McGurk effect, which is "a perceptual phenomenon that demonstrates an interaction between hearing and vision in speech perception" [8]. Finally, supporting multimodal redundancy, it was found that "The use of several modalities for processing exactly the same information provides an increased bandwidth of information transfer." [9].

According to J.Bouchet et al., a potential risk with using multimodal redundancy is that it may provoke cognitive overload and extra articulatory synchronization problems [10]. This is an important consideration, and this led to more research into the cognitive processes that happen during multimodal interaction.

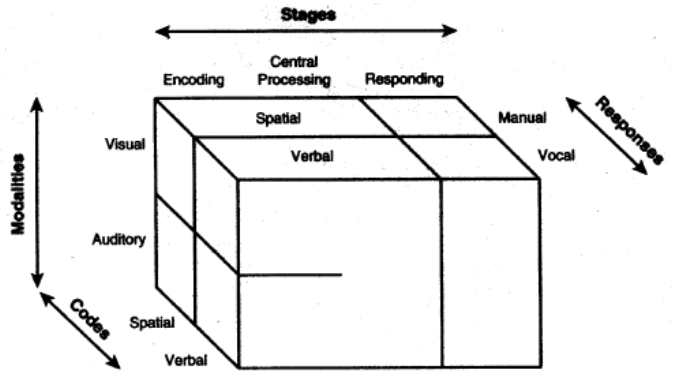

Figure 1: Wicken's Multiple Resource Model

The Wicken's Multiple Resource Model explains which resources are used when processing multisensory information that's coming into the brain (see Figure 1). According to the model, information reaches the brain through modalities such as visual and auditory input. That input is then coded as being spatial or verbal and is stored in working memory. When a person is responding to something, different resources are used than if the person is encoding or processing information. The resources needed are segmented into different cells in the diagram. Working memory has a finite capacity, and if a cell of the model is filled up, adding new information to that cell will cause an overloading problem. Some key findings are that the auditory and visual modalities do not share the same resource space, and neither do spatial and verbal information, meaning that they can be processed simultaneously without any problems [11].

This is elaborated further by the Modality Effect: "The Modality Effect focuses on the mode (visual, auditory, or tactile) of incoming information and states that we

process incoming information in different modes using separate sensory systems. Information is not only perceived in different modes, but is also stored separately; the Contiguity Effect states that the simultaneous presentation of information in multiple modes supports learning by helping to construct connections between the modes' different storage areas." [11].

From a cognitive perspective, the benefits of a multimodal come with several conditions. The simultaneous presentation of multiple input modes is good for memory, whereas sequential presentation is actually bad for memory [11]. Therefore, multimodal systems need to ensure that system output modalities are well synchronized temporally [7]. People can only deal with a finite amount of information at once, so a system should not overload people with too much incoming sensory information [11]. As mentioned in J.Bouchet et al.'s concern about extra synchronization problems, some processing capacity must be allocated to combining the information streams of multiple incoming modalities [11]. Finally, as a general rule of thumb, additional modalities should only be added to the system if they improve satisfaction or efficiency [7].

Discussion

From this literature review, it can be concluded that there are substantial benefits to using a system with multiple redundant modalities, however the system needs to be designed in a way that avoids cognitive overload. This can be achieved by properly following Wicken's model, to make sure that the individual resources needed to process each modality do not overlap. An unanswered factor is finding out how much mental focus is needed to synchronize incoming streams of information.

The overall potential benefits of the proposed multimodal reading system seem high, especially considering the effectiveness of the Orton-Gillingham Approach and the compelling McGurk effect.

Conclusion

This study aimed to answer the question of whether a multimodal reading system would be beneficial. Such a system was found to have more benefits than drawbacks, and this was shown from previous research related to multimodal redundancy, as well from related projects. Since the project has been found to be theoretically beneficial, the next steps are to study the system's technical feasibility, and to eventually build a working prototype.

This study brought up several more questions to investigate further. In the proposed system, each word could be accompanied by either an abstract sound or speech synthesis of the spoken word. In the apple example, using the spoken word "apple" might have the most obvious benefit, but if the abstract sound for apple was the same in every language, this might have another kind of benefit. If abstract sounds were used, then an algorithm would generate the sound of each word. Another big question is to determine what kind of sounds would be the most appropriate to generate.

References

1. Orton-Gillingham. Wikipedia. https://en.wikipedia.org/wiki/orton-gillingham.
2. Multisensory Activities to Teach Reading Skills. 2015. https://mnliteracy.org/sites/default/files/multisensory_techniques_to_teach_reading_skills.pdf.
3. Tammet, D. Different ways of knowing. 2011. https://www.ted.com/talks/daniel_tammet_different_ways_of_knowing?language=en.
4. Biedert, R. Text 2.0. https://text20.net.
5. Modality. Wikipedia. https://en.wikipedia.org/wiki/modality.
6. Edwards, A. Multimodality in Language and Speech Systems. Multimodal interaction and people with disabilities, (1999). https://www.speech.kth.se/milass/.
7. Skantze, G. Multimodal Interaction & Interfaces. 2015.
8. McGurk effect. Wikipedia. https://en.wikipedia.org/wiki/mcgurk_effect.
9. Multimodal interaction. Wikipedia. https://en.wikipedia.org/wiki/multimodal_interaction.
10. Bouchet, J., Madani, L., Nigay, L., Oriat, C., and Parissis, I. Formal Testing of Multimodal Interactive Systems. Engineering Interactive Systems Lecture Notes in Computer Science, (2008), 36–52.
11. Herrli, G. Multimodal Perception: When Multitasking Works. A List Apart, 2015. http://alistapart.com/article/multimodal-perception-when-multitasking-works.