Introduction

In this study we want to evaluate a newly designed input technique for writing text on a computer with a keyboard. The novel presents the user with the option to autocomplete written words after writing the first few letters. We hope that this technique will reduce the time needed when writing text. For comparing writing speed we measured how many words per minute the participants were able to write with and without the novel technique. To this end we presented various sentences to the participants, which they had to write as fast as possible.

Therefore we came to the following hypotheses:

H0: Our novel input technique does not affect writing time, when writing a sentence.

Application and Test Design

For testing autocompletion, we wrote a pyqt application which utilizes a custom completer based on the QCompleter class[1]. In the application, the user is presented with an instruction text and a text input field. After writing the first two characters of a word in the input field, a small pop up window will be shown at the bottom right of the current text cursor position. The pop up contains a top to bottom list of words with possible autocompletions, based on his current prefix (previously written characters of the current word). The user can then use the arrow keys to choose between the presented autocompletions and then press enter to autocomplete. He can also press escape to close the autocompletion pop up window. Additionally the user can also choose an autocompletion by clicking on it with the left mouse button[2].

The presented autocompletions are taken from a text file. The text can be a list of words or any other type of text like a dictionary for example. If the text file contains duplicates of words, they will be removed. The only real condition for using a text file for autocompletions is that each word has to be separated by either a whitespace or a new line.

The instruction text above the text input field contains an instruction which reads "please write the shown sentence(s) below" and the sentence the user is supposed to write. The sentences shown are taken from another text file which should contain a list of sentences separated by a new line. The sentences are shown in the same order as written in the text file, where the next sentence is presented after the user confirms his previous input by pressing the "enter" button on his keyboard[3]. The application closes when the input for the last sentence is confirmed.

All test data was logged to csv files. We differentiated between two log types, the first is the stats log which contains all variables for a task, like its completion time, and the second one is the event log which logs all events like keystrokes, or autocompletion of a word. For our analysis we only used the stats log, but the event log could be used for a deeper analysis of specific test iterations.

---

[1] Overall design based on: https://doc.qt.io/qt-5/qtwidgets-tools-customcompleter-example.html
[2] the repository also includes a demo video showcasing the application
[3] The "enter" button is not registered if the user is currently choosing an autocompletion

Participants

Due to limited resources and the current situation we were only able to test two participants who were also the same people who designed this test. They are both male, between 20 and 30 years old and are proficient in working with a keyboard and computer. As the test is designed as an in between study each of the participants were tested once with, and once without the novel technique . As we only had 2 participants, but wanted to gather more test data, each participant repeated the experiment an additional time so all in all, both were tested two times for each condition.
Due to the small number of participants and the fact that they took part in designing the test our results can not really be seen as meaningful. Instead this test can serve as an example of how similar studies could be conducted.

Test Setup

For each test we presented the participants with three sentences. They were instructed to write each sentence as fast as possible. Each participant completed the same test twice, once with and once without the autocompletion feature. For each written sentence we calculated the words per minute by dividing the number of words in a sentence by the task time in minutes and then the average word length including whitespaces. The task time is defined by the time it took from the first keyboard input to confirming the completion of the sentence by pressing "enter". As such, the participants were able to read the sentence for as long as they liked before they started writing without affecting their task time.

Variables

Independent variables: order of the tests, use of novel technique (on/off)

Dependent variables: words per minute

Controlled variables: input device (laptop keyboard), screen resolution, screen size environment, autocompletion words, presented sentences.

Possible confounding variables: learning effect, fatigue

To reduce the effect of the confounding variables of fatigue and learning effect for testing if the novel technique affects writing speed we implemented counter balancing. This means that half of the participants were tested with the novel technique first and without the technique second and the other half were tested the other way around. As the tested sentences stay the same and in the same order, learning effect

Visualisation

Words per minute per user with and without novel technique
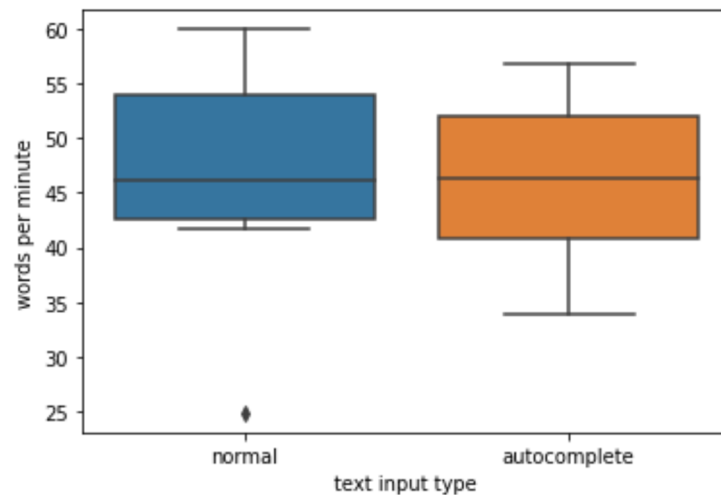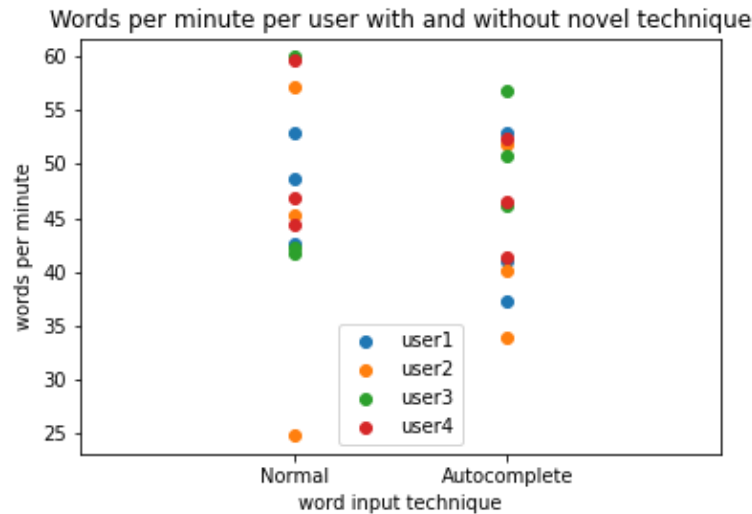
Discussion

The statistical test results show that the difference in words per minute between the normal text entry (M = 45.92 SD = 6.88) and our improved technique, the auto-completion, (M = 47.19 SD = 9.31) is not significant, with the t-test results: t=0.405, p=0.69.
Therefore we cannot reject the H0-Hypothesis that our novel input technique does not affect writing time, when writing a sentence.
While the boxplots show that the words per minute were higher for the normal text entry, the difference is only marginal. Comparing the different users also shows that some participants were faster with the auto complete method and others without. All in alls the results are not very conclusive.

Limitations

It is probable that the test is highly influenced by the possible words for autocompletion. More words also mean that the user has to choose between more possibilities of autocompletions (which means more inputs if the desired word is further down the list of

presented completions) or that he has to input more characters before he chooses a completion. For our test we included all words of our chosen sentences, some similar words for each word as well as a few random sentences. In a real world example there are different examples how the word lists could be filled like using an entire dictionary, words the user has written previously or a list of words or text, specific to the use case of the application. For a possible follow up study, the used word list should be defined by clear standards and/ or a specific use case.

Additionally for our test we only tested a few sentences, as we only had very few participants and not enough time for long experiments. This makes our results less meaningful as it is unclear if our results would be the same for other sentences as the length of words and the sentences themselves will probably affect the words per minute. A follow up study with more participants and a larger time frame for testing would benefit from having a larger selection of sentences. It would also be better to choose sentences at random, instead of having the same sentences for each participant[4].

Our biggest limitation is our very small sample of participants and the fact that they were the same people who designed the test. As such both participants knew which sentences they had to write beforehand, and also had extensive knowledge of the test itself. It is highly unlikely that this did not influence our results, and as such, our results can not be seen as conclusive.

Conclusion

In conclusion we could not find any significant differences in writing time with and without our autocompletion technique. As we had some very strong limitations, these results can not really be seen as meaningful. To get conclusive results, this study should be repeated with some small changes and most importantly a much larger set of participants, who had no part in designing the experiment.

---

[4] As our dependent variable is words per minute, differing sentences can still be compared