Comparing the Typing Performance of a Soft-Keyboard with and without Haptic-Feedback

***Abstract*—A huge amount of text is today written on the touch screen of a smart phone or tablet, especially text messages. These devices usually implement a haptic feedback through very short vibration. This feedback is even the default on most phones today. Still there is very little research on whether it actually has an impact on the typing performance. Therefore, we conducted an experiment on the effect of the haptic feedback on the typing performance. We found the haptic feedback to be error reducing. Because errors usually need to be corrected, we concluded that it might influence the overall typing performance on a touch screen keyboard.**

I. Introduction

In today's mobile age nearly everybody owns at least some sort of smart device with a touch screen. The usage of touch screens in products has become very popular among the manufactures, because it is a relatively cheap way of allowing the users to interact with a certain device. Products ranging from smart phones to even household appliances already include touch input and the number of devices is growing fast. This technological change brings forward some new challenges concerning the way of user input on such devices. Before everything changed to touch related input, simple buttons or keyboards with direct haptic feedback were used. Since a screen is just a plane of glass no feedback is provided to the user by default. Therefore, vibration motors are used in some cases (smart phones and tablet) to simulate a button being press. Especially for text input where the human is used to feedback from a physical PC keyboard this lack of feedback might have an influence on the typing performance. There are a lot of other methods like swiping, speech to text and handwriting recognition, but many users still prefer the standard keyboard input. Some older studies [1] [2] imply that there is no significant benefit in performance when using haptic feedback simulation. We think that these statements do not hold up to the modern touchscreen and feedback technologies of the smart phone era. Therefore, we want to reevaluate this research question and check whether there are significant advantages of using haptic feedback for text input on a smart phone.

II. Related Work

We started our research by looking into older literature to get some inspiration for our own study, as well as evaluate some of results of prior studies.

The topic of haptic feedback of a soft keyboard has been discussed in very few papers and studies. The researcher focus rather on different keyboard layouts [2] or even different hardware keyboards [3]. Castellucci et al. researched various types of feedback, but in the end they did not get any significant results. They stated that the type of feedback does not have a significant influence on the speed or the accuracy of the text input [1]. Similar results were obtained by Boehm et al. [3]. They did compare a physical and a soft keyboard on similar smart phones and did not see and benefit in using the hardware keyboard. It is important to note that the validity of the study of Boehm et al. might be questionable, since the participants were not used to the hardware keyboard on the smart phone. The question arising from this is: do the older statements about soft key boards still hold true or did the spreading of soft input methods (adapting to soft input) change the behavior of the people.

We got a lot of ideas from other papers regarding the measurement of the performance and usability aspect. For example Roeber et al. used metrics like entry speed, error rate and user satisfaction to evaluate a projection based keyboard [4]. The speed factor was measured by the words per minute achieved by the participant and additionally the satisfaction level was measured on a scale from 1 to 5, by asking for the participants opinion on the different systems. Another attempt to measure similar metrics of performance was proposed by Azenkot et al. [5]. They also measured the words per minute, but in a different way. Every five completed characters were counted as one word, even if it was not a word that would make sense. The mean error for every phrase they evaluated was as their error rate. In a questionnaire they determined the participants prior experience of using smart phone keyboards [5]. Exact formulas to calculate different metrics for error rate and entry speed were proposed by Kim et al. [6]. Since the users were not forced to type correctly, they used in addition to the total error rate the corrected and uncorrected error rate. To evaluate and compare their input methods they used an analysis of variance (ANOVA). Castellucci et al. gathered text entry metrics on android devices to compare different input methods [1]. They compared the methods of typing, handwriting recognition and swiping for input. For their evaluation, they used the ANOVA test method with the standard metric words per minute and error rate.

In many of the papers and studies the same data set was used. For comparing text entry techniques MacKenzie et al.

Independent	Dependent
Haptic Feedback Type	Task Completion Time
	Total Error Rate
	Subjective Opinion

TABLE I
DEPENDENT AND INDEPENDENT VARIABLES OF OUR STUDY.

generated a special phrase data set containing 500 items [7]. The phrases are in English and have no punctuation, which is a time critical factor on mobile devices. The setup of many studies were quite familiar. One field where the text source is shown and a second field below it for the user input. This type of layout was used by MacKenzi et al. and Castellucci et al [2] [8].
We used the gathered insides into these studies to derive our own design and methods.

III. Research Question and Hypothesis

In our empirical project we want to evaluate the impact of haptic feedback on the performance of typing with a soft keyboard on a smart phone. We wanted to only focus on typing words letter by letter on a basic touch keyboard. No additional help or other techniques such as swipe or auto correction would be available during testing. Additionally, we wanted to focus on a relatively young participant group (20 to 30), since those people are very familiar and well trained smart phone users. The main research question of our project was: Does the usage of haptic force feedback improve your typing performance on a smart phone?
With the research question outlined, we concluded our hypotheses. Since we had a idea about what the result might look like we decided to use an on-way approach for evaluation. Therefore, we set the following hypothesis:

- H_0 : Using haptic force feedback for text input on a smart phone performs equivalent or even worse than using no feedback at all.
- H_1 : Haptic force feedback will improve the overall performance of typing on a smart phone soft keyboard.

Based on these hypotheses we began to prepare the study design in the next step.

IV. Data Collection

After setting the hypotheses we had to outline the variables and testing procedure.
Since we wanted to measure the performance of the keyboard input and performance is a latent variable, we had to use operationalization in order to obtain variables that we could measure. The dependent and independent variables are listed in Table I. We want to vary the haptic feedback of the soft keyboard on the Android device in two stages (default vs. none) and measure the error rate and speed of typing. The speed will be calculated by measuring the time the participants need to type a predefined text divided by number of words, which gives us the word per minute metric. We use the corrected error rate as mentioned by Castellucci et al. [1], since we force the user to correct his mistakes before proceeding. All values will be collected by our custom Android App. Additionally, a questionnaire must be filled out by the participants at the end of the study. The questions will include general information as well as some specific details about the usability of the two tested methods.
Since we want to collect our data in a field experiment, there might be many confounding variables we have to account for. Possible examples include:

- Personal smart phone settings
- Skillset of the user
- Learning on a new device
- External environmental influences

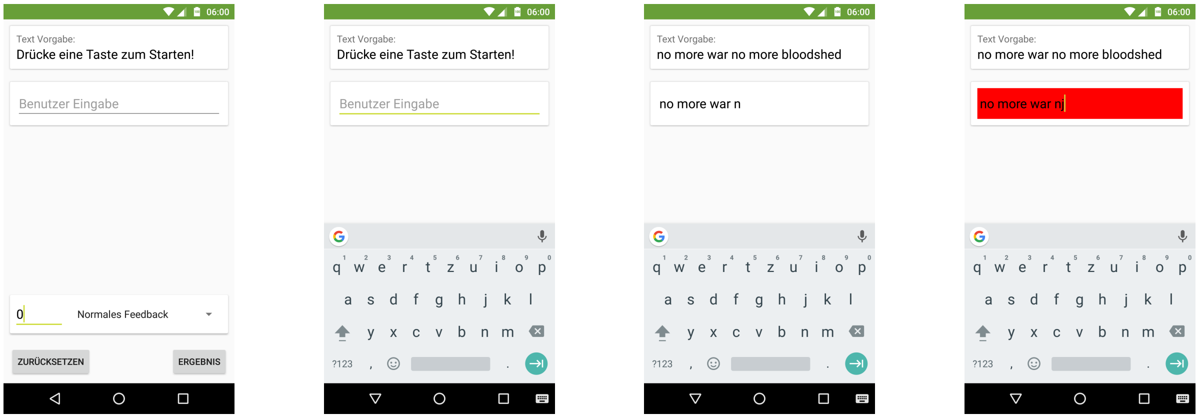
To avoid problems and reduce negative effects as much as possible we tried to retain most confounding variables constant. For example personal settings or other influences such as notifications or device size were avoided by using the same test smart phone for all tests. Furthermore, we varied some other confounding variables (15 phrases picked from 500 and test pattern) randomly to ensure there are little to no dependencies between the tests.
For simplicity we decided to use within subjects testing. This would allow us to gather more data in shorter period of time and ensure that the data would be comparable within one test. In total, we intended to test with a relatively large group of students (around 30) to ensure even distributed data and valid results.
Concerning the internal and external validity of the study, our goal was to find a good balance. On the one hand we control and manage a lot of the testing process (measurements, smart phone type and settings), but on the other hand we conduct the study in the field and not a laboratory. Of course there might be some trade offs, but a good balance between these two validity factors was more important to us.
Another critical point of any user study is the overall quality consisting of the reliability and objectivity. Since we control all the measurements very precisely through our custom Android App, we can assume that our reliability is keep very high. We also ensure a good objectivity through our good documentation of the test steps.

V. Experiment Preparation

In order to record the previously described variables, we settled on implementing a custom Android Application. This App would allow us to track the exact time, user input characters and error rate of each individual participant. Furthermore, we developed and additional questionnaire to gather some general information about the participants, as well as some subjective feedback about the conducted tests.

A. Android App

We decided to implement our own custom Android App, since we had prior knowledge in App development and we did not find a suitable ready to use alternative. The basic layout idea was inspired by the user study design of the previously



(a) Default start screen with all the settings.

(b) Participant view with the the label settings hidden behind the keyboard.

(c) Typical typing scenario with source text on top and user input on the bottom.

(d) Error state identifying that the user input does not match the source.

Fig. 1. Overview of the basic Android App screens and their functions.

mentioned related work. There are just three important graphical user interface (GUI) elements:

1) Source Text Field: Shows the source sentence the user has to enter.
2) User Text Input Field: Normal text field showing the current input of the user. Background will turn red if the input and source do not match.
3) Soft Keyboard: Standard Google Android keyboard with default settings. Only the haptic feedback setting gets varied.

The described layout and the different states are illustrated in Figure 1. Additionally to the three study relevant GUI elements, further input fields and buttons (unique user id and current feedback mode) for labeling the recordings were added. This information would be entered by the instructor before and in between each of the two individual tests. During the test these irrelevant data is hidden behind the keyboard.

To start the test the user had to input a random character on the keyboard. Afterwards, the first source sentence was shown and the timer was started. As a source we used the English MacKenzie "Phrase Set for Evaluating Text Entry Techniques" [7]. From there on, the user had to copy the text character for character and as fast and accurate as possible. If the source and the user input did not match by one or more, the error state counter was increased by one and the color of the input text field changed to red, indicating an error state. Only the entering of the error state was counted and not the individual character based errors. As an example: If I would enter two wrong characters at a time the error counter would only increase by one. Only if I corrected the mistake and then make another one then it would count up again. The user had to correct the input before proceeding. After one entire phrase was entered the time, error, user input and source were saved and the next sentences was shown right away. In the end, after completing the 15 sentences all the results were saved in a JSON file with the time stamp and user id as file name.

B. *Questionnaire*

The second step of data gathering was a very short and basic questionnaire. This would allow us to compare the recorded data from the Android App against the perception of the participants. Each questionnaire included the user id and the test order, which were filled out by the instructor beforehand. The rest of the data was entered by the participants through Google Forms in a browser on the test smart phone after the two typing tests were conducted. First of all, some questions about age, profession and personal smart phone settings were asked. Afterwards, the speed and accuracy of each of the two tests should be judged on a likert scale from one to six. In the end, the user should also select their favorite method of the two tested ones. All the data from the questionnaire was stored in a CSV file for later processing.

VI. Experiment Conduction

The study was carried out over a few days with no fixed schedule. Random student participants were selected by approaching them in their free time in the University of Passau. If they were interested and wanted to participate we started the field experiment at their current location.

At first the instructor prepared the Android App and the questionnaire by filling out the user id and the test pattern. Afterwards, the task at hand (not revealing the research question) was explained to the participant by showing how to start the test and demonstrating the case of a wrong input. If no questions were asked, the student had about one to two minutes to get familiar with the test device and do some practice runs before hand. The keyboard settings during the learning phase matched the first test pattern. Therefore, it was also varied during the course of the entire study. After the demo session possible questions about the usage of the app were answered and the app was reset. The first of the two tests was started. For each test a total number of 15 random phrases from the MacKenzie data set had to be entered. When the participant

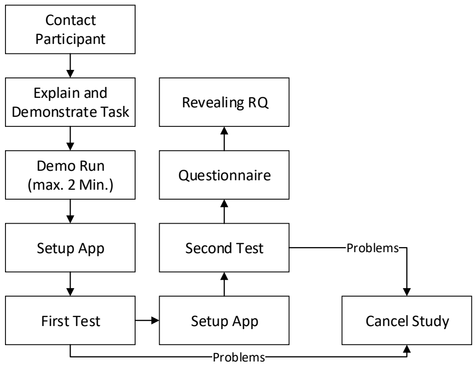


Fig. 2. Test procedure diagram, illustrating the individual steps of the study.

was done with the first part, the settings were changed and the test repeated with the other test pattern. In the end, the user was asked to fill out the Google Forms questionnaire. Afterwards, the data was stored in a Dropbox folder and in a short discussion with the participant the intentions of the study were revealed. Any severe problems during the study (device issues or extreme external interference) would have caused a abortion of the test run and the data would have been discarded. This did not happen during our testing. The entire testing procedure is illustrated in Figure 2.

VII. Data Evaluation

After we gathered all the data from our 26 participants we had to review and evaluate the recorded data set. Since the information from the App was saved in individual JSON files and we wanted to work with IBM SPSS, we had to convert it into on large CSV file for importing.

A. Participants

In total 26 participants (19 male, 7 female) ageing from 16 to 27 years old took part in our study. The mean age was 21.65 years with a standard deviation of 2.66. Most of them were students at the Univerity of Passau (21 persons), the other 5 participants were either still in school or finished University. The 26 users mostly use Android (65.4%) smart phones personally and only around a third (34.6%) use iOS devices personally. When asked about their personal keyboard settings in the questionnaire many of the participants had to double check by testing it on their own phone. Around 69.2% use some sort of haptic feedback as a default setting, while the other 30.8% preferred no haptic feedback at all. Other settings like audio feedback (19.2%) or swipe text input (19.2%) were only used by a few people. Regarding the input hand, most of the people (73.1%) use two hand to type text. The others (26.9%) preferred typing only with one hand. Another question was about the experience with the usage of a smart phone in general. The answers ranged from 2 (rather inexperienced) to 6 (expert) with a median of 4 and a mean of 3.92. This indicates that the participants judged themselves as above average users on the 1 to 6 scale.

	Shaprio-Wilk Sig.	Paired T-Test
WPM (no Feedback)	0.966	0.025
WPM (def. Feedback)	0.597	
Errors (no Feedback)	0.447	0.003
Errors (def. Feedback)	0.201	

TABLE II
RESULTS OF THE SHAPIRO-WILK TEST AND PAIRED T-TEST. THE DATA IS NORMALLY DISTRIBUTED AND SHOWS A SIGNIFICANCE BETWEEN THE TWO INPUT METHODS.

B. Android Application Data

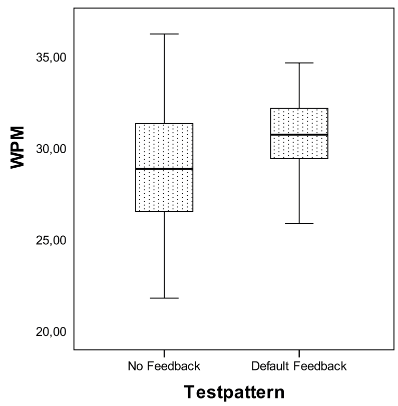
Just by taking a look at the box plots in Figure 3, one can observe that there is a difference between the two tested methods. With a mean of 30.66 words per minute (WPM) the participants were faster with the haptic feedback compared to the 28.76 WPM without any feedback. In addition, less typing errors were made (mean with feedback: 10.96, mean without feedback: 16.62) with the default haptic feedback enabled. Another important observation is the standard deviation. Both measured parameters show a greater deviation when no feedback was used.
For the evaluation of the data from the Android Application we opted for a simple T-Test since we had metric values and only two test factors. Since we had the intuition that using haptic feedback would increase the words per minute metric and reduce the overall number of errors we selected the one-sided T-Test with paired samples. The paired samples are important, because we used within subject testing. Before analysing the significance of the data set, a Shapiro Wilk Test was used to determine if the data for each test setting was normally distributed. The results are shown in Table II. All of the significance levels of the Shaprio Wilk Test are greater than 5% indicating that the data follows a normal distribution. The paired T-Test (shown in Table II) revealed that there is a significant difference between the two tested methods. Both significance levels (2.5% and 0.3%) are below the limit of 5%.

C. Test Questionnaire

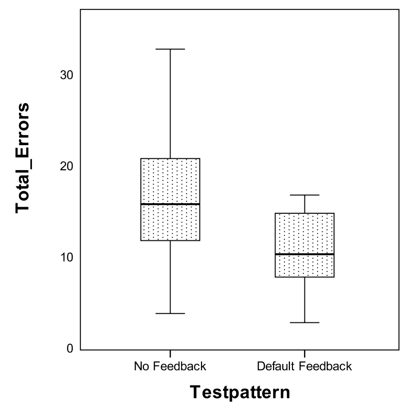
The second data source was the questionnaire the participants had to fill out after the study. The interesting metric here are the subjective judgement of the speed and accuracy, as well as the preferred method. From the data in Table III it was already pretty clear that the difference between the two input methods might be not significant different. The Man-Whitney-U test, which we used because of the ordinal data from the questionnaire, underlined this observation. The standard deviation of the speed and precision data is always larger for the no feedback test. To the question which method the participants preferred 76.9% said that they liked the haptic feedback more, but only 23.1% stated that no feedback was their favorite method during the test.

VIII. Interpretation

From the study data of the Android App it is pretty clear that there was a significant benefit of using haptic feedback on the soft keyboard. This finding was underlined by preferred



(a) Box plot illustrating the words per minute metric for both tested input methods.



(b) Box plot of the total error count for both default and no haptic feedback tests.

Fig. 3. Boxplots of the two variables measured by the Android App.

	Mean	Median	Man-Whitney-U
Speed (no Feedback)	3.92	4	0.455
Speed (def. Feedback)	4.19	4	
Precision (no Feedback)	3.38	4	0.059
Precision (def. Feedback)	4.03	4	

TABLE III

MEAN AND MEDIAN OF THE SPEED AND PRECISION RATING IN THE QUESTIONNAIRE, AS WELL AS THE RESULT OF THE MAN-WHITNEY-U TEST TO DETERMINE SIGNIFICANCE BETWEEN THE TWO METHODS.

method of input, which was selected in the questionnaire. But on the other hand the subjective test revealed that the participants did not really notice a significant difference between the used methods. Both the speed and accuracy questions reveal somewhat the same results, even though there was a difference measured with the app. Another question one might ask: If the error rate of no feedback is higher and the user have to correct the errors, would the speed not be affected by the number of wrong inputs? This is a thought that came to our mind after analysing the data. It is pretty obvious that a higher correction rate will lead to less words per minute in the test. This thought could relate to the fact that the significance level of the subjective precision is only slightly above the threshold of 5%, but the speed judgement is far off with 45%. The participants noticed that they made more mistakes without feedback, but they did not account for the time they needed to correct them, when asked for the typing speed.

Referring to the research question and hypotheses, the test data supports the proposed H_1. Therefore, we accept H_1 for our special user study. But in general, the hypotheses regarding the performance can not be completely accepted or rejected. On the one side there is a speed difference in the App data set, but this might be only the case for this certain testing method with error correction. On thing that was shown in the study is the increase of the error rate when no feedback was used.

IX. CONCLUSION

In conclusion, this user study gave a small insight into the usage of haptic feedback in combination with smart phone keyboards. The results showed that using haptic feedback is faster and users made less errors with it. But there might be a connection between the number of errors and the words per minute. Other researchers conducted similar studies and have shown that there is no speed benefit. Out study showed a difference in speed, but this could be explained by the corrections that had to be made. Overall we can at least conclude that using no haptic feedback increases the number of errors made when typing on a soft keyboard. In order to get an better insight of the connection between speed and the higher error rate and if there is really a difference, further more elaborate studies need to be conducted. This was just an exploitative field test to get a general idea about the topic.

REFERENCES

[1] S. J. Castellucci and I. S. MacKenzie, "Touch me, hear me, feel me: Feedback preferences for mobile text entry," in *Proceedings of the International Conference on Multimedia and Human Computer Interaction–MHCI 2013*, 2013, pp. 123–1.

[2] I. S. MacKenzie and S. X. Zhang, "The design and evaluation of a high-performance soft keyboard," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 1999, pp. 25–31.

[3] P. Bhm, A. Hansbauer, and C. Wolff, "Sind richtige tasten besser? vergleich touch- bzw. tasten-basierter texteingabe auf dem smartphone," in *Mensch und Computer 2016 - Tagungsband*, W. Prinz, J. Borchers, and M. Jarke, Eds. Aachen: Gesellschaft fr Informatik e.V., 2016.

[4] H. Roeber, J. Bacus, and C. Tomasi, "Typing in thin air: the canesta projection keyboard-a new method of interaction with electronic devices," in *CHI'03 extended abstracts on Human factors in computing systems*. ACM, 2003, pp. 712–713.

[5] S. Azenkot and S. Zhai, "Touch behavior with different postures on soft smartphone keyboards," in *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*. ACM, 2012, pp. 251–260.

[6] J. R. Kim and H. Z. Tan, "A study of touch typing performance with keyclick feedback," in *Haptics Symposium (HAPTICS), 2014 IEEE*. IEEE, 2014, pp. 227–233.

[7] I. S. MacKenzie and R. W. Soukoreff, "Phrase sets for evaluating text entry techniques," in *CHI'03 extended abstracts on Human factors in computing systems*. ACM, 2003, pp. 754–755.

[8] S. J. Castellucci and I. S. MacKenzie, "Gathering text entry metrics on android devices," in *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2011, pp. 1507–1512.