EFFECTS OF AUDITORY FEEDBACK ON MULTITAP TEXT INPUT USING STANDARD TELEPHONE KEYPAD

Sami Ronkainen

Nokia Mobile Phones
User Centric Technologies Laboratory
P.O.Box 50, FIN-90571 Oulu, Finland
sami.ronkainen@nokia.com

Juha Marila

Nokia Research Center
Visual Communications Laboratory
P.O.Box 407, FIN-00045 Nokia Group, Finland
juha.marila@nokia.com

ABSTRACT

Text input in mobile devices is becoming even more important with the increasing amount of text-based services such as mobile email. The keypad input of small mobile devices could benefit from well designed auditory feedback. In this paper three different key feedback sound schemes are studied in multitap input of a standard telephone keypad [1]. In this paper is the effect of three different feedback schemes to the efficiency of text input were studied. Also the user preferences during a two-week usage period were observed.

1. INTRODUCTION

The standard keypad used e.g. in mobile phones consists of twelve keys: numbers 0-9, the "*" key and the "#" key. Text input using these keys is implemented so that each key (except keys "*", "#" and "1") represents several letters of the alphabet. Traditionally this has been implemented so that e.g. for typing the letter "q" the user has to press the number 7 key two times. There is a time limit during which the second one of the subsequent key presses must happen. During that time the device will interpret the user's behaviour so that he is using the multiple key presses for selecting one of the letters represented by that key. After the timeout has happened, the next key press is interpreted as a new letter. In other words, to input letters "q" and "p" in a row the user must first press the number 7 key two times fast enough that the delay between the key presses is shorter than the predefined timeout. Then he needs to wait for the timeout to pass. Then he can press the number 7 key once again, to input the letter "p".

In many manufacturers' mobile phones the typing is made faster by allowing the user to kill the timeout by pressing a key not representing any letters - such as an arrow key - before the timeout has passed, in order to accept the currently presented letter. The alternative meaning for that key can in that case be ignored as the key press is interpreted only as killing the timeout. In the previous example the user could have pressed the number 7 key twice, then press the arrow key to indicate that the letter "q" that is currently selected, is the one that he wants. Then he could immediately have gone on to pressing the key 7 once more to input the letter "p".

The problem in killing the timeout is that the keys not representing any letters, usually have some other meaning. If it then happens that the user presses the alternative key in order to kill the timeout, and it happens that the timeout just passed by itself, the alternative usage for the key will take place. For example if the alternative key is an arrow key, it may happen that the cursor moves without the user noticing it. The result, when typing quickly, may be that the user will input a few more letters before he notices that he has moved the cursor in error.

In this paper a new concept of indicating the multitap timeout with a distinctive tone sequence is presented. Its effect on typing speed and errors were tested. Also two different lengths of regular key press tones (with no multitap timeout tone indication) are compared. Two separate tests were arranged. One compared the key tone concepts initially when all the test users were unfamiliar to the new concepts. Another test was arranged after a two-week usage period, measuring the effect of learning and performance in a divided-attention usage situation. Also subjective preferences towards the new concepts were studied.

The hypothesis in the research was that the multitap timeout indication tone will inform the users that the timeout happened, enabling them to go on typing the next letter faster than before. From existing research it was also known that reaction times to simple auditory stimuli are faster than to simple visual stimuli [2].

Obviously this would apply only to the users who do wait for the timeouts instead of killing them. It is a fair assumption that the performance of this type of users will reflect also performance of novice users. Usually people having bought their first mobile phone do not notice the possibility of timeout-killing immediately when starting to type texts with it.

If the user has a habit of killing the timeouts, the error situations where the user is trying to kill the timeout by pressing an alternative key while the timeout has actually passed already should be noticed by the user, leading to fewer errors when typing. In divided-attention usage situations (such as typing when walking) it was hypothesized that auditory feedback would help the users to keep track of what they are typing and when they can type further letters, leading to fewer errors and potentially faster typing.

The counterhypothesis was that adding extra tones to the key input will cause more mental processing in the users' minds, leading to slower typing speeds. If that is the case, having no key feedback tones at all should lead to fastest typing speeds.

Yet another counterhypothesis was that the dominant aspect affecting the typing efficiency is not related to the cursor reappearing, but e.g. to the pace at which the users can move their fingers and find new keys. In that case none of the different sound feedback aspects should be better than others and it is the subjective preference of users that will define the best feedback scheme.

The key usability aspects to be measured in the study were the speed and efficiency of typing, as described in equations 1-3, and users' subjective preference towards different sound schemes.

2. TEST

The test equipment for the test were Nokia 6210 mobile phones. Four different tone schemes to be tested are listed in Table 1.

0	Silent
A	100ms key tone, no multitap timeout indication
B	8ms key tone, no multitap timeout indication
C	8ms key tone with multitap timeout indication consisting of two 6ms tones

Table 1. *Tested key feedback sound schemes.*

It was measured that the average time between when the phone no longer interprets the next key press as a part of a multitap character (and when the multitap timeout indication tone was played in C concept) and the time when the visible cursor appears on screen was 465ms. So when writing e.g. a 50-character message where e.g. 15% of characters require multitapping, the C concept could theoretically lead to a 3.5s improvement in typing speed.

Two tests were arranged. The tests consisted of the users typing short text messages using phones equipped with different auditory feedback schemes. In the initial test the schemes were compared to each other in an undistracted usage context. For finding out the possible differences in real life situations, the users were given phones equipped with the C auditory feedback scheme to be used for a two week period. After the usage period another test was arranged, consisting of two parts. In the first part learning of the concept was tested in an undistracted usage context. The second part tested users' performance in a usage context requiring divided attention. This was to simulate real life conditions where the users cannot fully concentrate on the typing task. Such situations are e.g. typing while walking or having a conversation with another person.

There were 12 test persons altogether, all of which were Nokia employees. As the reason for the test was to study the auditory feedback schemes, it was decided to utilize users already familiar with the other aspects of the phones and used to typing text messages using the phone keypad.

2.1. Test method

The users were presented with a computer screen showing the text to be typed one sentence at a time. The sentences selected were familiar Finnish proverbs, such as "Hädässä ystävä tunnetaan" ("A friend in need is a friend indeed"). The reasonnel for choosing proverbs as the texts to be typed was to avoid errors due to people misreading or having to learn the texts. Proverbs were already familiar to them so they could concentrate on the typing instead. The proverbs were selected so that they would include more multitap characters than Finnish language in general, but because they also had to be familiar, this goal could not be completely met.

The sentences were presented in groups of six. Between each group the test user switched the phone to be tested. The order of auditory feedback schemes presented was selectively varied so that as many permutations as possible were used. That was done to avoid effects resulting from learning, or user fatigue in the test, or user getting more acquainted with the test situation.

To help later analysis, and to avoid having to use even bigger number of permutations, the texts were normalized. In each block of six sentences, the total number of characters, and the number of characters requiring entry from the same key (i.e. leading to having to wait multitap timeouts to occur) were the same. The number of characters requiring multitapping (as e.g. the letter "o", compared to the letter "m") was not normalized, as it was assumed that in natural language they would be evenly distributed among sentences. Also the distances between the successive key presses (e.g. moving the finger from button 2 to button 3, compared to moving from button 2 to button 8) were not normalized in the texts, for the same reason.

The time spent typing each sentence, as well as the number of corrected and remaining errors, was measured from a digital video tape. As it was sometimes difficult to say at which point exactly the user pressed a key to type a letter, all times measured were rounded to the closest half a second.

Before each test the users were given a chance to familiarize themselves with the test by typing a test sentence.

The users were instructed not to correct any errors except for the ones they notice right after making the error. This was to avoid spending time moving the cursor around. They were also instructed that if the cursor moves accidentally, they can leave it where it is and continue typing. They were also instructed that if they notice a word missing they can type it anywhere in the message.

The rationale behind the instructions was to measure the actual time spent typing – not spent on finding and correcting errors. But it is very natural to instinctively correct the error right after making it that the users were not forced to avoid their natural behaviour. The time spent on error correction was cleaned out from the final times in the way shown in equation 1.

There were two key measure for the users' performance in the test. *Efficiency* is the success rate in the typing, divided by the time spent typing, as shown in equation 1. The success rate is calculated by subtracting the percentage of typing errors from 100%. Both the amount of errors corrected while typing and the errors remaining in the final result were taken in account in calculating the error percentage.

The time spent typing was cleaned up for efficiency calculation so that the time spent on error correction was not included. This was done by calculating the average typing speed in the text to be typed - as shown in equation 2 - and multiplied with two times the amount of corrected errors. The multiplication with two was justified so that when the user makes an error, he first has to press the Clear key to remove the erroneous character, then type the correct one. So effectively he has typed two extra characters.

$$Eff = \frac{100 - (100 * \frac{RErr + CErr}{text\_len})}{time - cps * 2 * CErr} \quad (1)$$

Where

RErr is the amount of errors remaining in the text after typing,
CErr is the amount of errors corrected while typing, and

$$cps = \frac{text\_len + 2 * CErr}{time} \quad (2)$$

In other words, the amount of characters per second is the actual amount of characters (the characters in the text plus the amount of extra characters typed due to error correction) divided by the actual time spent typing.

The typing speed by itself was measured by *words-per-minute* (WPM), as defined in equation 3.

$$WPM = \frac{text\_len + 2 * CErr}{60 * time * 5.5} \qquad (3)$$

Where

text_len is the actual length of the typed text,
time is the actual time spent typing,
CErr is as above.

The number 60 comes from that the times were measured in seconds. The number 5.5 was used as the average length of a word. As the WPM data were only compared to each other, it is insignificant which number to use - 5.5 is a commonly used number for word length.

In the second test a divided-attention task was presented. It introduced an additional computer screen the users were instructed to look from the corner of their eye while typing. The screen changed colours in one-second intervals. The colours were chosen randomly from red, black, white and different shades of blue. Additionally, whenever the screen changed colour, also a digit (0-9) was presented. The users were instructed to keep track of what is happening on the screen, and whenever they notice a red colour, or an even digit (or both) occurring, they must react by using a foot switch. If they did not react in three seconds, a “punishment” was presented in a form of a fairly annoying sound. This was introduced in order to motivate the users really to concentrate also on the secondary task.

The test setup was created by using the Presentation (http://www.neurobehavioralsystems.com/software/presentation/) software.

2.2. Initial Test

In the initial test the four different sound feedback schemes were tested against each other. The purpose was to find if there were any significant differences between the schemes.

It was found in the test that 7 out of 12 test users had the habit of killing the timeouts by using arrow keys. As the hypothesis said, for them the typing speed should not be affected but only the error rates. The types of users were named to "wait" type users and "kill" type users.

It was also found out that none of the test users used keypad tones at all in their current mobile phones.

As the sentences were arranged in groups of six, and there were four auditory feedback schemes to be tested, each user typed 24 sentences.

The test results can be seen in Table 2 and Figures 1 and 2.

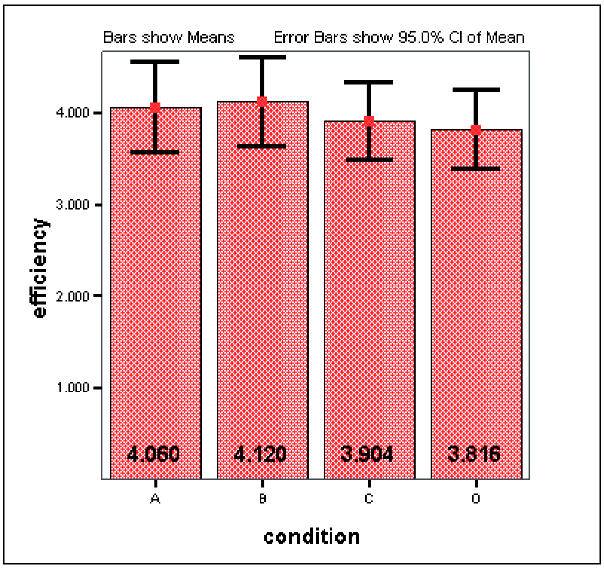

Figure 1. *Average efficiency of different feedback schemes (all users)*

The average efficiencies for each sound concept can be seen in Figure 1. The efficiencies were compared to each other using Student's t-test. The differences did not prove statistically significant. So it can be said that none of the concepts was better than another with respect to typing errors.

It must be pointed out, however, that the amount of errors due to the users trying to kill the timeout after it already has passed was very small. So that type of errors hardly contributed to the overall error rate at all. Obviously avoiding that kind of errors did not lead to differencies in error rates among the concepts either.

The same was found when measuring typing speed. The WPM results are presented in Figure 2. The t-test proved that also there no statistically significant differences could be found.

If the multitap timeout tone would bring extra benefit to typing speed of those users who waited for the timeouts to happen, it should be visible in the WPM numbers. However, there was no strong evidence that this would be the case. The fastest typing on the average did indeed occur with the C concept. The difference, however, was not statistically significant, even when compared to the 0 concept where slowest typing occurred. If there were such benefit, it should have shown especially in the difference between B and C concepts where the only difference is the timeout tone. But the average WPM numbers of the "wait" type users was 9.9 for the B concept and 10.2 for the C, as shown in Table 2.

	0	A	B	C
"wait" WPM	9.5	9.8	9.9	10.2
"wait" eff.	3.28	3.28	3.32	3.50
"kill" WPM	11.9	12.3	12.8	12.5
"kill" eff.	4.36	4.75	4.82	4.56

Table 2. *Average WPM and efficiency for different sound feedback schemes and different types of users*

The differences in WPM numbers between "wait" and "kill" types of users were statistically significant ($p<0.01$) in all concepts, as could be expected.

As said, no significant differences could be found between the sound concepts. However, the users' subjective preference was favourable towards the C scheme. 7 out of 12 users gave positive comments towards the multitap timeout indication tone. Some of the users who killed the timeouts pointed out that the timeout indication tone is useless for them, and commented only on the differences in the actual keypad tones.

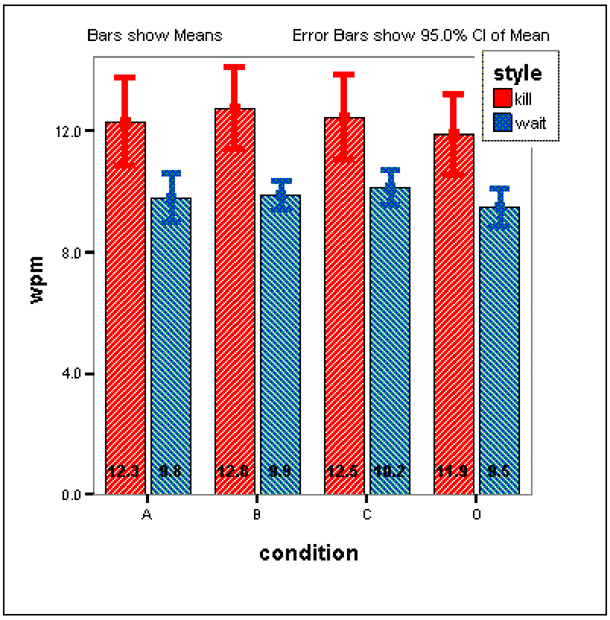

Figure 2. *Average WPM in 1st test by user style*

2.3. Final Test

After the first test each user was given a phone equipped with the C sound feedback scheme for a two-week usage period (working weeks, i.e. 12 days). Two users dropped out of the test because they felt they could not cope without their regular phone. So finally there were 10 users who were tested after the usage period.

During the usage the users were asked to keep the key tones on whenever possible and to otherwise use their phone just like they use their regular one. Every day they were queried about how many text messages they had sent during the day. On the average they sent 2.98 messages per day. Assuming that an average message would be 50 characters, this means that they typed about 1800 characters during the test. That equals to typing about 64 messages of the same length (28 characters) that were used in the laboratory tests.

As the first test didn't reveal any significant differences between the sound feedback schemes, only two sound schemes were tested in the first part of the second test. The silent (0) condition was tested against the C condition. The purpose was to find out if learning and using the sound feedback would cause any improvements in the typing speed or efficiency.

The t-test showed a statistically significant difference between 1st and 2nd test silent condition, in WPM rate , as shown in Table 3.

test	**condition**	**wpm**	**t (1 vs 2)**	***p***	**df**
test 1	0 - silent	10.94			
test 2	0 - silent	12.06	-1.89	**0.06**	118
test 1	C - tone	11.54			
test 2	C - tone	12.32	-1.11	ns	118

Table 3. *Mean WPM rates of silent and C tone conditions in both tests*

There is no significant difference in the second test between silent (12.06 WPM) and tone (12.32 WPM) conditions, however. So learning did occur, but it was related to the overall usage of the phone - its physical measurements, placement of the keys etc. Not to the actual sound concepts.

As mentioned, also usage context requiring divided attention between the phone and some other task was simulated. To find out the pure effect of the multitap timeout tone, the sound schemes tested were 0, B and C in the divided-attention task. The results can be seen in Table 4.

	0	B	C
"wait" WPM	6.9	7.4	7.4
"wait" eff.	2.39	2.32	2.27
"kill" WPM	11.0	10.7	10.9
"kill" eff.	3.53	3.52	3.65

Table 4. *Average WPM and efficiency for different sound feedback schemes and different types of users in the divided attention task*

In the divided attention task there was a hint that having some sound feedback at all could have speeded up the typing of the "wait" type users. The t-test ($p=0.11$) did not show a statistically significant difference when comparing e.g. scheme C to scheme 0 but it can nevertheless be seen as an indication that such difference might exist. This was also supported by the comments of some users. They said that it was easier to type when they heard how many times they had pressed the key already.

Interestingly there was hardly any difference in writing speed between B and C sound schemes even with "wait" type of users. For the "kill" type of users it was natural that no such differences exist since they don't get to hear the multitap timeout tone in any case. The divided attention was real, however. Comparing the users' performance in the divided attention task using either sound scheme (0 or C) to the same schemes in the focused attention situation lead to statistically significant ($p<0.01$) differences.

When looking at the efficiency figures to find an effect of different sound feedback schemes to error rates, no significant differences could be found from either user group using any sound feedback scheme.

However, the subjective feedback after the two-week period was even more positive than after the initial test. 8 out of 10 users said they would use it in their own phone. Even some of the "kill" type users said they had felt the multitap timeout tone to be beneficial for them. So in real life the tone would seem to be more useful than was shown in the tests. Some of the "kill" type users commented that there are e.g. situations requiring single-hand usage where timeout killing is difficult.

As for key tones in general most of the users said they would not switch silence to the tested keypad tone. 4 out of 10 users said spontaneously that they would like to have the phone with the timeout tone only, i.e. without any other keypad sounds.

3. DISCUSSION

As shown, there were hardly any measurable differences between the different sound feedback schemes. The only finding that was somewhat backed up by the results was that having auditory feedback from the keypad would seem to speed up the typing of "wait" type users in a divided-attention usage situation. So it can be said that in general keypad tones do not seem to have any significant effect on typing speed or efficiency.

On the subjective level there are differences. In the tested user group, keypad tones in general were disliked. But there were very positive remarks towards the C concept. This indicates that also differences in performance may still exist in some level, even though they did not show in the test.

One possibility is that giving the feedback through two modalities lessens the cognitive load required when typing. It may be that the users perform almost equally well with or without sound feedback but have to concentrate more if they have to rely on visual feedback only. Of course this phenomenon should presumably have been visible from the results of the divided attention task. But it can be that the task was still not demanding enough, or that the typing speed and efficiency are not the best measures for the phenomenon.

Another possibility is that the benefit from the timeout tone is only in the users' minds. Perhaps when the tone occurs, they first hear the tone, then see the cursor and feel that the tone happened first so it must be benefitial.

Some users commented that they learn roughly the length of the timeout anyway so they get only minor benefit from the timeout tone. This would seem like a plausible explanation for the lack of measurable differences between the C concept and the other ones. As mentioned, theoretically there could have been a big improvement. This could be verified by arranging a test where both visual and auditory feedback would be removed and the users would just have to learn the length of the timeout.

In all it can be said that the differences are subtle, if there are any. Choosing texts with even more multitap characters could have produced more differences but at the risk of introducing typing delays and errors due to users having to type awkward texts. In natural language, of course, the benefit in typing speed gained from the timeout tone would be even smaller.

Also the error situation where the timeout tone would most probably be benefitial - i.e. a "kill" type user moving the cursor by accident - is fairly rare. So its effect is not visible in the error rates in the test even though in real life such errors do happen occasionally.

4. REFERENCES

[1] -, "Human Factors (HF); Assignment of alphabetic letters to digits on standard telephone keypad arrays", ETS 300 640, European Telecommunications Standards Institute, August 1996

[2] M.S. Sanders, E.J. McCormick, *Human Factors in Engineering and Design, p286*. New York: McGraw-Hill, 1993.