Comparison of Two Text Entry Methods on Interactive TV

Adam J. Sporka Ondřej Poláček Pavel Slavík
Faculty of Electrical Engineering, Czech Technical University in Prague,
Karlovo nám. 13, 12135 Praha 2, Czech Republic
{sporkaa, polacond, slavik}@fel.cvut.cz

ABSTRACT
This paper describes a study with 18 participants on two text input methods for use with an interactive television, The Numpad Typer (TNT) and TwiceTap. Both methods were designed for use with a common infra-red remote control. Each character is entered using two keystrokes. After one hour and forty minutes of practice, split into 5 sessions, the participants were able to type faster than 9 words per minute. While the performance of both methods was similar, the TwiceTap required a smaller number of keystrokes per character, was less prone to errors, and was found to be more comfortable by the participants. Over time, the use of frequent *n*-grams in TwiceTap was reinforced while the use of infrequent *n*-grams was suppressed.

Categories and Subject Descriptors
H.5.2 Information Interfaces and Presentation: User Interfaces—Input devices and strategies

General Terms
Design, Experimentation, Human Factors, Measurement, Performance.

Keywords
Text input, Interactive Digital TV, performance, remote control, user study.

1. INTRODUCTION
Interactive Digital Television (IDTV) is a growing technology in terms of the number of users. According to the Leichtmann Research Group, 30% of all households in the US in 2011 have at least one TV set connected to the Internet . The IDTV enhances the television with interactive content, including web browsing and searching, social media, e-mail, etc. One of the vital parts of these applications is the text input.

There has been an extensive research in this context on various devices and methods of interaction, such as use of gestures, tablets, or mobile telephones. However, commercially available products are still significantly more expensive than the infra-red remote controls used commonly today. In this paper a user study of two text input methods with 18 users is presented. The methods can be operated using a common infra-red TV remote.

2. RELATED WORK
Probably the most popular text entry method on keyboards with a limited number of keys is Multi-tap, commonly available on the mobile telephones. Each key is assigned three or four characters. To enter a character the user has to repeatedly press the same key until the desired character appears. The cursor advances upon a predefined timeout (user's inactivity). According to MacKenzie et al. [9] the KSPC measure (keystrokes per character) is 2.2. There were several modifications reducing the KSPC: Less-Tap (=1.6) [12], where the sequence of letters on the keyboard was no longer alphabetical, or LetterWise (=1.15) [10] where the layout of keys dynamically shifted according to the previously entered text, which was on the expense of the clarity of the layout.

The T9 by Tegic Communications [4] is another method intended for use on mobile telephones that can be implemented on a TV remote control. The alphabet is subdivided in the same way as in the Multi-tap. The user selects the desired characters by pressing the corresponding keys and after a sequence of keys is entered, the word is disambiguated using a dictionary. Even though T9 is almost optimal in terms of the KSPC (close to 1), many users find it too difficult to use [5].

One of the first work on a text input method for the IDTV was Immargson's *The Numpad Typer* (TNT) [7] (see Figure 1). All characters can be entered by two keystrokes on a 9-button keypad on the remote control. The characters are laid out in two levels of 3×3 grids. The first keystroke selects a group. The second keystroke selects a character in that group. The speed of the keyboard varied from 9.3 up to 17.7 WPM (words per minute). A similar entry method, in which characters are entered by two keystrokes, was patented by Kandogan et al. [8]. This method is designed for a 12-button keypad commonly used on mobile phones. There is no evaluation of this method. Twist&Tap [1] is designed for text input on a TV using a remote control with accelerometers. Vega-Oliveros et al. [13] explored a multimodal text input on IDTV combining speech input, Multi-tap, and a virtual keyboard.

Iatrino et al. [6] compared Multi-tap and a virtual QWERTY keyboard for text input on IDTV. The virtual QWERTY keyboard was a grid reminiscent of a standard PC keyboard. Arrow keys were used to select a character. The “OK” button confirmed the selection. In a study with 36 participants, Multi-tap outperformed the virtual QWERTY in terms of speed and user satisfaction. However, a similar study by Geleijnse et al. [3] indicated that the virtual QWERTY keyboard is preferred by the users. The type rate of the virtual QWERTY, Multi-tap, and T9 are similar.

3. Introducing TwiceTap
We developed a novel method of text input called TwiceTap. The method is similar to the work by Kandogan [8]. It has the same layout of characters as the Multi-tap. Each character is entered by a sequence of two keystrokes. See Figure 2 for an example: To enter the letter `t`, first the key *`t`*`-u-v` is pressed to select this group. Individual letters of this group are then

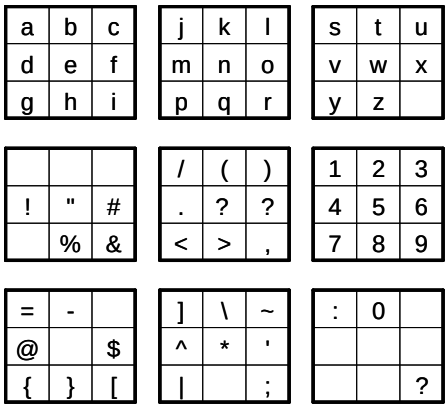


Figure 1: Layout of TNT

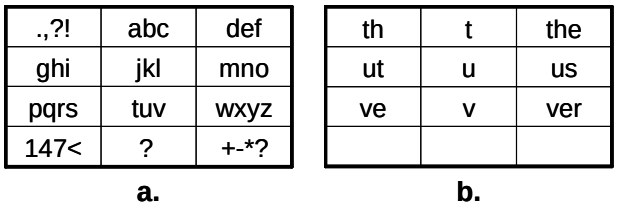


Figure 2: Layout of TwiceTap – a) as shown before the first keystroke, b) as shown after the user hit key tuv

displayed on separate keys in the same column (Figure 2b). The letter is entered upon pressing the corresponding key. Unlike the Kandogan's method, the blank spaces were filled with the two most frequent *n*-grams for each letter. The term *n*-gram further refers to a sequence of *n* characters. The *n*-grams with the length equal to one, two and three character are being called unigrams, bigrams and trigrams respectively. There is a fixed set of *n*-grams which were selected using a corpus containing 50,000 most common English words. The space, as the most common character in English, can be entered using one keystroke only.

4. Experiment

In this paper we perform a comparative study of the TwiceTap method and the TNT method [7]. Both methods can be implemented using low-cost hardware.

The Multi-tap method would seem like a logical choice of the baseline, however we decided against it. We indeed intended to compare our method with a baseline that is established in the HCI literature but at the same time we were seeking for an input method that was unknown to the users so that the skills to use each method would be balanced (the TNT met these requirements) while the Multi-tap is a method that is known and used by a majority of the mobile telephone users.

18 participants (10 men, 8 female, mean age = 22.7, SD = 2.49) took part in a controlled experiment. They were not familiar with text input on IDTV, however they were all experienced mobile phone users. (3 of them used T9 on their phone; 2 people used a QWERTY keyboard. The rest of the group used Multi-tap.)

We expected KSPC=2 for the TNT, while for the TwiceTap the KSPC should be less than 2 because of the single-keystroke space and the availability of *n*-grams. A 9-button keypad is required for the TNT and a 12-button keypad for the TwiceTap. Both methods use a reference chart on the TV screen. The TNT occupies about 4 times more space on the screen than the TwiceTap.

The experiment was organized in 10 sessions per participant. Each session lasted 20 minutes. The participants were allowed to complete up to two sessions per day and used each method for 5 consecutive sessions. The order of the methods was counterbalanced in order to minimize any learning effects.

A 40" Samsung LCD TV was used for both methods. A remote control shipped with the Motorola VIP1003 set-top box was used. The participants were seated on a sofa about 2 m in front of the screen. We implemented an audio feedback after pressing a key for both methods. Their task was to copy given phrases from Mackenzie's phrase set for evaluating text entry methods [11] as fast and accurate as possible. The use of *n*-grams in TwiceTap was suggested but not enforced. A high-pitched click sounds upon the first keystroke, a low-pitched click upon the second, and a buzzer sound upon an error. The participants had to correct any typing errors before the application let them continue. At the end of the experiment the participants were asked to complete a post-test questionnaire to evaluate both methods. The participants were given a reasonable financial compensation for their time at the end of the experiment.

Our group of 18 users entered total of 118,811 characters of text. 57,048 characters (48%) were typed using TNT, 61,763 characters (52%) using the TwiceTap. The participans' type rate, measured in WPM, improved with each session for both methods, eventually reaching 9.1 WPM (TNT) and 9.6 WPM (TwiceTap). The type rate did not reach a plateau by Session 5, which indicates that the type rate reachable by the experts would be even higher. (See Fig. 3)

KSPC for TwiceTap was lower than for the TNT. In Session 5 it was about 2.25 for TNT (the users had to press at least two keys per one character on average) and 1.75 for TwiceTap. A paired *t*-test of the KSPC values measured for individual participants on the Session 5 confirmed that the difference was significant: $t(28.58) = 16.3$; $p < 0.001$. The TwiceTap method therefore requires less amount of physical effort than the TNT. The values somewhat decreased over the course of the entire data collection, by 5% (TNT) and 8% (TwiceTap) due to the decreasing number of errors. The error rate has been calculated as the number of backspaces pressed by the participants per amount of the correctly entered characters. Its development over the sessions is shown in Fig. 4. The mean error rate over all participants of the TNT and TwiceTap was about 5.1% and 3.8% in the last session. A paired *t*-test indicates that the TwiceTap was significantly less prone to errors than the TNT ($t(26.55) = 2.31$; $p < 0.05$).

A number of words could be written in different ways that were mutually exclusive. E.g. the word `"version"` could be entered using any of these sequences of elements: `(ve)(r)(s)(i)(on)`, `(v)(er)(s)(i)(on)`, or even `(ver)(s)(i)(on)`. We defined a usage of an element (an *n*-gram available at the keyboard or an individual character) as the ratio [number of occurrences in text] / [number of actual instances of use]. A KSPC of less than 2 in case of the TwiceTap implies that the participants made a systematic use of the *n*-grams. The Fig. 5 shows the combined usage of the most and the least frequent *n*-grams, compared against the usage of all *n*-grams. The overall strategy of the *n*-gram use has apparently shifted over the sessions: While the use of the most frequent *n*-grams was reinforced (60% in Session 1; 70% in Session 5), the usage of the least frequent *n*-grams dropped by 12% (for bottom 10 *n*-grams) resp. 23% (bottom 5).

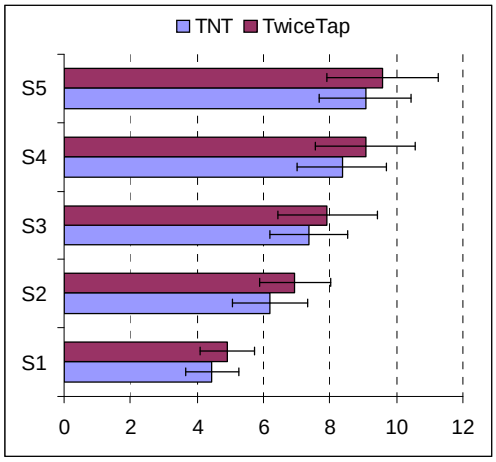


Figure 3: WPM by method and session. Mean values over all participants with SD in error bars for sessions S1 through S5

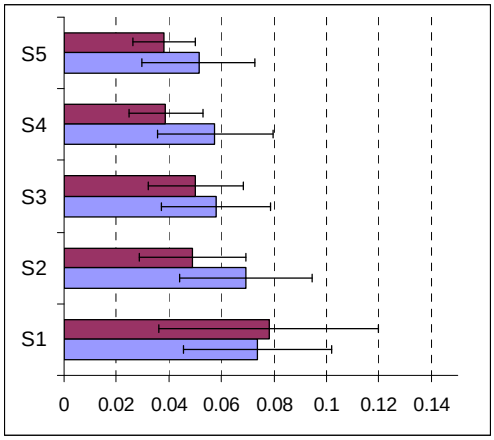


Figure 4: Error rate by method and session. Mean values over all participants with SD in error bars

This suggests that people recognized that the utility of the *n*-grams varied with their frequency, and learned to use those that were most frequent while giving up on those that did not notably contribute to the overall performance. (In the last session, there were total of 1322 opportunities to use one of the top 5 *n*-grams while only 193 opportunities to use one of the bottom 5 *n*-grams.)

The average time per character (the time needed to type a single letter of the English alphabet) was 1325 ms using TNT and 1633 ms using TwiceTap in the last session. A paired *t*-test of the times per character for TNT and TwiceTap shows that the TwiceTap was significantly slower in terms of typing individual letters ($t(49.44) = -4.396$; $p < 0.001$). This may be explained by the fact that the users did not see the whole TwiceTap layout at once. The final position of the letter was revealed only after the pressing of the first button, thus forcing the users to perform the visual search twice. The letter distribution in the test phrases roughly observed the common letter distribution in the English language, which in decreasing frequency was: `e`, `o`, `a`, `t`, `s`, …, `v`, `j`, `x`, `q`, `z`. It could be expected that the more frequent a letter, the shorter time per character. The actual time per character, however, did not observe this expectation, as shown in Figure 6. Detailed

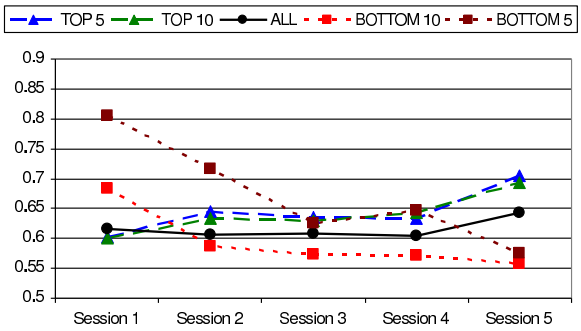
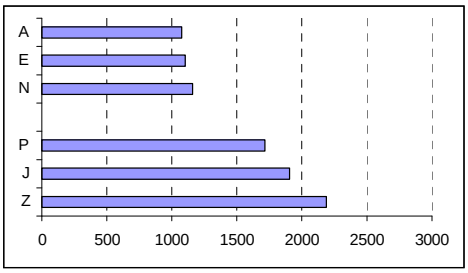
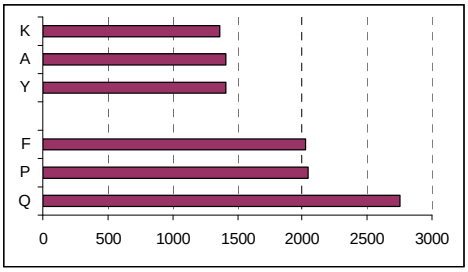


Figure 5: Usage of the most and least used *n*-grams by all users per session.



a)



b)

Figure 6: Average time in milliseconds (horiz. axis) to enter individual characters (vert. axis). Three "fastest" and three "slowest letters" reported. a) TNT, b) TwiceTap

analysis suggested that the further apart the keys needed to enter a particular letter, the longer time per character.

For subjective evaluation we asked participants to complete a post-test and System Usability Scale (SUS; [2]) questionnaires for each method. They were also asked to make comments on both methods. The SUS was designed to quantify subjective usability of a product to allow comparisons. The item no. 5 ("I found the various functions in this system were well integrated") was not used in our case. A paired *t*-test was conducted to compare the SUS scores. There was a significant difference in the SUS scores for TNT (Mean=56.0, SD=20.8) and TwiceTap (Mean=70.7, SD=11.8); $t(17) = -2:58$; $p < 0.05$: Usability of TwiceTap method was therefore perceived significantly better than TNT. In the post-test questionnaire, five questions for each method were used to quantify other aspects such as fatigue and comparison with mobile phone text input methods. Five-scale Likert items were used (1="certainly no" ... 5="certainly yes") for the following questions as listed in Table 1. Median scores are displayed in Table 2. A paired Wilcox test showed that scores for the TwiceTap in Q1 and Q3 were significantly better than for the

TNT. We also asked the participants which method is better to use. 11 participants preferred TwiceTap, 4 preferred TNT, 3 expressed no preference.

Participants commented mostly on the keyboard layout and visual presentation of both methods. The two keys in the TNT were too far apart for some frequent letters (4 participants gave such a feedback), which caused pain in the thumb (1 participant) and a need to check position of fingers on remote control often (1). The TwiceTap layout is similar to Multi-tap, which was found positive (7) or negative (1). The *n*-grams in the TwiceTap were found useful (8). The TNT displayed too many letters on the screen (2), however visibility of all letters was appreciated (4); "I can search for the next letter while entering current letter" (2). The special characters such as apostrophe or exclamation mark were difficult to find on the TNT (2), but easy on the TwiceTap (1).

Table 1: Likert items for the post-test **questionnaire**

Q1	Method was comfortable to use.
Q2	Method was physically demanding.
Q3	Method is better than Multi-tap.
Q4	Method is better than T9.
Q5	I would like to use method on my TV.

Table 2: Median scores for post-test **questionnaire and p-value for each item. Lines set in bold report a significant difference.**

Question	TNT	TwiceTap	p
Q1	**3.5**	**4**	**0.0113**
Q2	2.5	2	0.0617
Q3	**2**	**3.5**	**0.0077**
Q4	3	3	0.2885
Q5	3	4	0.1324

5. Conclusion

This paper presented a comparative study of two text input methods for Interactive Digital Television—the TNT and the TwiceTap. While the TNT supports only letter-by-letter typing, the TwiceTap also allows typing of the most frequent *n*-grams. The letters in both methods are entered by two keystrokes on a common TV remote control. Users preferred the TwiceTap, mostly because of its similarity to Multi-tap and the availability of *n*-grams. Even though the type rate was similar for both methods (TNT: 9.1 WPM; TwiceTap: 9.6 WPM), the keystrokes-per-character rate was significantly lower for the TwiceTap, indicating that this method required less physical effort. The TwiceTap was less prone to errors than the TNT.

Over the course of the experiment the participants adapted their own *n*-gram use strategy. By the end of the experiment they increased the use of the *n*-grams that were most frequent in the English language while ceasing the use of the less frequent ones, preferring to enter individual characters instead.

The TwiceTap keyboard can be further optimized by shuffling the position of characters on each key according to the layout of Less-Tap [12]. Other interesting extension would be dynamic *n*-gram suggestion based on the already written text.

Acknowledgements. This research has been supported by the MSMT research program MSM6840770014

6. REFERENCES

[1] Aoki, R., Maeda, A., Watanabe, T., Kobayashi, M., and Abe, M. Twist&tap: text entry for TV remotes using easy-to-learn wrist motion and key operation. Consumer Electronics, IEEE Transactions on 56(1) (February 2010), 161–168.

[2] Brooke, J. SUS: A Quick and Dirty Usability Scale. In Usability Evaluation in Industry, P. Jordan, B. Thomas, I. McClelland, and B. Weerdmeester, Eds. Taylor & Francis, London, 1996, 189–194.

[3] Geleijnse, G., Aliakseyeu, D., and Sarroukh, E. Comparing text entry methods for interactive television applications. In Proc. of the 7th European conference on European interactive television conference, EuroITV '09, ACM (New York, NY, USA, 2009), 145–148.

[4] Grover, D. L., King, M. T., and Kushler, C. A. Reduced keyboard disambiguating computer. Tech. rep., US Patent Publication, 1998.

[5] Gutowitz, H. Barriers to adoption of dictionary-based tex*t*-entry methods: a field study. In Proc. of the 2003 EACL Workshop on Language Modeling for Text Entry Methods, TextEntry '03, Association for Comp. Linguistics (Stroudsburg, PA, USA, 2003), 33–41.

[6] Iatrino, A., and Modeo, S. Text editing in digital terrestrial television: a comparison of three interfaces. In Proceedings of EuroITV'06 (Athens, Greece, 2006).

[7] Ingmarsson, M., Dinka, D., and Zhai, S. Tnt: a numeric keypad based text input method. In Proc. of the SIGCHI conference on Human factors in comp. sys., CHI '04, ACM (New York, NY, USA, 2004), 639–646.

[8] Kandogan, E., and Zhai, S. Two-key input per character text entry apparatus and method. Tech. rep., US Patent Publication, 2004.

[9] MacKenzie, I. KSPC (Keystrokes per Character) as a Characteristic of Text Entry Techniques. In Human Computer Interaction with Mobile Devices, F. Paterno, Ed., vol. 2411 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2002, 405–416.

[10] MacKenzie, I. S., Kober, H., Smith, D., Jones, T., and Skepner, E. Letterwise: prefix-based disambiguation for mobile text input. In Proceedings of the 14th annual ACM symposium on User interface software and technology, UIST '01, ACM (New York, NY, USA, 2001), 111–120.

[11] MacKenzie, I. S., and Soukoreff, R. W. Phrase sets for evaluating text entry techniques. In CHI '03 ext. abstr. on Human factors in computing systems, CHI EA '03, ACM (New York, NY, USA, 2003), 754–755.

[12] Pavlovych, A., and Stürzlinger, W. Less-tap: A fast and easy-to-learn text input technique for phones. In Graphics Interface (2003), 97–104.

[13] Vega-Oliveros, D. A., Pedrosa, D. d. C., Pimentel, M. d. G. C., and de Mattos Fortes, R. P. An approach based on multiple text input modes for interactive digital tv applications. In Proceedings of the 28th ACM International Conference on Design of Communication, SIGDOC '10, ACM, New York, NY, USA, 2010, 191–198.