Optimized and Hygienic Touch Screen Keyboard for Large Letter Set Languages

Asad Habib[1,2]
[1]Institute of Information Technology, Kohat University of Science and Technology, Pakistan.
Telephone number +92-300-5951911
habib-a@is.naist.jp

Masakazu Iwatate
HDE, Inc.16-28, Nanpeidai, Shibuya, Tokyo
150-0036, Japan.
masakazu.iwatate@hde.co.jp

Masayuki Asahara
National Institute of Japanese Language and Linguistics,
Center for Corpus Development
10-2 Midori, Tachikawa, Tokyo
190-8561, Japan.
masayu-a@ninjal.ac.jp

Yuji Matsumoto
[2]Graduate School of Information Sc.
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara
630-0192, Japan.
matsu@is.naist.jp

Wajeeha Khalil
Department of CS and IT
University of Engg. and Technology, Peshawar, Pakistan.
Telephone number +92-91-9218423
wajeeha.khalil@nwfpuet.edu.pk

ABSTRACT

1) Does QWERTY type keyboard provide optimum performance for every language? 2) Is it suitable for touch-screen devices?
In this paper we tried to find out answers to these questions. QWERTY keyboard remained the most common mode of input for a long time. The recent growth of touch screen devices is forcing changes in modern GUIs (Graphical User Interfaces). They stipulate designing new input systems with higher performance, more user friendliness and better usability etc. The common QWERTY type keyboards lack these qualities when deployed on smaller touch screen devices such as smart phones, e-readers, tablet PCs and PDA (Personal Digital Assistant) etc.

Categories and Subject Descriptors

H.5 [**INFORMATION INTERFACES AND PRESENTATION**]: Graphical user interfaces (GUI) – *Evaluation/methodology, Ergonomics, Input devices and strategies (touchscreen).*

General Terms

Performance, Design, Human Factors, Standardization, Languages.

Keywords

Touch screen keypads, Non-Roman input systems, Urdu input method editor, Hygienic input systems design, Perso-Arabic script input.

Touch screen devices have assorted form and figures having non-intersecting platforms. Typing non-Roman text on modern touch screen devices is hard, particularly languages with complex orthography and letter sets larger than English e.g. Urdu, Thai, Hindi and Kashmiri etc. Out of top 10 most influential languages of the world, 60% are non-Roman. Optimizing input systems for Chinese and Japanese etc. are much more difficult and involve extra steps. In this paper, we proposed novel and distinct keypad that has been optimized for precise, simple, quick and efficient typing of Urdu text.

It is equally appropriate for other Perso-Arabic script languages e.g. Arabic, Persian and Punjabi etc. With trifling code changes, our proposed keypad is useful for non-Perso-Arabic script languages with letter set larger than English e.g. Hindi, Kashmiri and Thai etc. For Romanized text input using English QWERTY keypad on touch screen, we recommend computer graphics techniques of 'onion skinning' and 'invisible buttons' to create adequately large and easily visible buttons on smaller touch screen devices. This reduces strain on the eyes of the user even when a system is used continuously for long hours. For optimization, character unigram and bigram frequency distributions are used to facilitate fast, correct and easy composing. High frequency letters are typed by stronger typing fingers. Alternate hand typing is also taken care of. We carefully designed our proposed keypad such that it offers better visibility, usability, extendibility, aesthetics and user friendliness.

We carried out evaluation of our systems in two ways; 1) automated procedures showed 36% improvement over the existing systems. 2) For real world performance testing, we conducted user evaluation of our proposed system and compared it with in-the-market generic QWERTY keypads for touch screen gadgets. The touch screen Acer Iconia tablet PC was used as test bed for user evaluation. Our proposed interface showed 21% improvement over the existing ones. Detailed results are presented in the subsequent sections.

1. INTRODUCTION

In line with the growth of touch screen devices, IMEs (Input Method Editor/Environment) and on-screen/virtual keyboards have been hot areas of research lately [1][15]. Typing native Urdu script on general touch screen gadgets is exhausting. Many modern gadgets either lack a good interface for typing Urdu e.g. Apple iPhone, or provide sluggish, inconvenient and hard to use keypads. There is no widely used agreed-upon keyboard or IME for Urdu [5].

We live in the spread out age of touch screen devices with expected growth in future. Currently available input systems were primarily developed for PCs. They need improvement in usability efficiency and visibility etc. An ignored area of research was that QWERTY type keypads are not fitting for text input of languages with letter-sets larger than English. This problem is more noticeable in data entry for non-Roman script languages such as Urdu etc. It is spoken by a very large population due to the large Diaspora of Indo-Pak subcontinent residents in different parts of the world. On the contrary, the use of Urdu is quite limited in the digital world including internet. Among others, one of the reasons is the difficulty in composing Urdu on modern computers particularly the touch screen devices. This problem gets more critical on small screen handheld gadgets [6]. Other non-Roman languages also suffer from poor input methods and input systems despite the fact that these languages are 60% of top 10 most influential languages of the world [9].

We developed a novel keypad for non-Roman language Urdu which is compliant with five golden principles of Ergonomics i.e. Performance, Ease, Aesthetics, Comfort and Safety. Our suggested keypad has been optimized for accurate, easy, speedy and efficient typing on small touch-screen handheld gadgets. We carefully designed our proposed keypad so that it offers better visibility, usability, aesthetics and user friendliness. Our optimization technique for arrangement of alphabets is extendable and fitting to other natural languages with large letter-set, in particular the Perso-Arabic script languages such as Sindhi, Kashmiri, Punjabi and Pashto etc.

For evaluation of our novel proposed keypad, we performed two types of evaluations; a) Automated evaluation and b) Users evaluation. Our automated experiments on a large Urdu corpus reveal almost 36% improvement over contemporary keypads available in the market. We also carried out real world analysis through user evaluation.

The results of our evaluation are discussed in much detail in Section 7. The rest of the paper is organized as follows. Section 2 illustrates numerous character level NLP (Natural Language Processing) applications. Section 3 discusses Urdu language. It explains important issues related to Urdu text input and the challenges to develop Urdu IME. Section 4 is about additional design parameters. The Urdu keypads currently in use and our proposed keypad are discussed in Section 5. Experiments, model and methodology are discussed in Section 6. Section 7 is about comparison and evaluation of the proposed keypad. Section 8 concludes the paper. Future directions are mentioned in Section 9.

2. Character-Level NLP Applications

NLP is also called CL (Computational Linguistics). It is a vast field of study that has numerous applications. These levels include inter sentential applications such as discourse analysis, sentence level applications and intra sentential applications e.g. phrase or words analysis etc. CL also deals with various applications at the "characters level" as shown in Figure 1. This research targets the area of 'development of IME, keypads and their Graphical User Interface Designs' etc. We have come up with novel keypad for text input on touch screen devices. Our proposed keypad is explained in detail in Section 5.

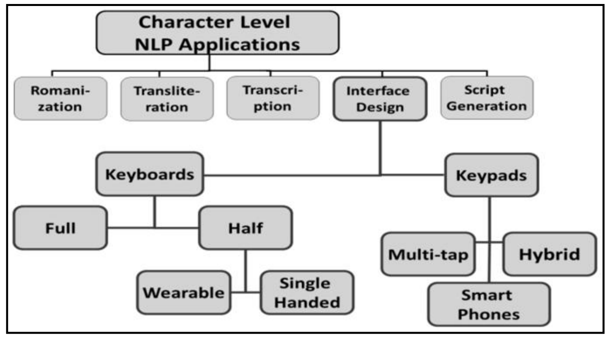

Fig. 1 Character level applications of NLP.

3. Urdu

Urdu is the Pakistan's national language and official language in few larger states of India e.g. UP (Uttar Pradesh). UP is India's most populous state. Urdu is the Lingua franca of Indo-Pak subcontinent and spoken in various parts of the world. This is mainly because of the large South Asian Diaspora around the world. Urdu has many interesting integral linguistic features such as grammar, orthography and rich morphology etc. Some salient features of Urdu language are mentioned in the following.

3.1 Rank of Urdu

Urdu is the national language of Pakistan. It belongs to the language family of central Indo-Aryan language [8]. It is spoken by a large population of speakers across a score of countries. Urdu is written from right to left in Perso-Arabic script. Its grammar is both gender and number sensitive. It is the 2nd largest Arabic script language according to the number of speakers [10][18].

Phonetically, Urdu is quite similar to Hindi. Written Urdu and Hindi use different and mutually exclusive scripts. Hindi is written in Devanagri script while Urdu is written in Perso-Arabic script. However, in spoken they appear to be the same language. The Ethnologue [19] considered Urdu and Hindi as the same language and ranked it the 5th largest language of the world according to the number of speakers. The numbers of Urdu and Hindi speakers are given by Table 1 [22].

Table 1: Hindi and Urdu Speakers (in millions)

	Native	*2nd Language*	*Total*
Urdu	60	104	164
Hindi	366	487	853
Total	426	591	**1,017**

3.2 Urdu Script

The term 'script' refers to the continuous natural and native way of writing Urdu text. Based on the correct and appropriate shapes of individual letters Urdu ligatures, words, phrases and sentences are formed. Collectively all of these are referred to as Urdu script.

Urdu is written from right to left. Arabic has 28 base letters while Persian has 32 letters. Both Arabic and Persian letter-sets are subsets of Urdu. However, the exact number of Urdu letters is not agreed upon. Numerous articles report different numbers of letters [4][14][22]. The largest letter set contains 58 letters introduced by NLA, Pakistan. It is shown in the following Figure 2.

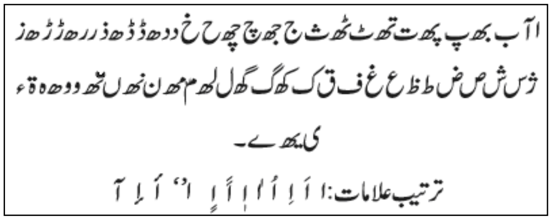

Fig. 2 The 58 letters-set of Urdu alphabets.

Designing optimized Urdu keypads for small screen widgets is a knotty problem. Relatively large letter-set and no agreement over the total number of letters in Urdu letter-set make the problem more complex. According to Afzal and Hussain Urdu alphabet has 57 letters and 15 diacritical marks [21]. Hussain [27] reported 41 letters in Urdu. Ijaz and Hussain [14] mentioned 56 letters. Habib et al. [4] reduced the Urdu letter-set to 38 basic letters as shown in Table 2.

Table 2: Basic 38 Urdu letters and their corresponding Roman letters for Romanization

Roman Letters	*Urdu Letters*	*No.*
a,e	ا	1
~	آ	2
b	ب	3
p	پ	4
T	ت	5
t	ٹ	6
S	ث	7
j	ج	8
C	چ	9
H	ح	10
K	خ	11
D	د	12
D	ڈ	13
Z	ذ	14
R	ر	15
R	ڑ	16
Z	ز	17
J	ژ	18
S	س	19
Sx	ش	20
Sx	ص	21
Zx	ض	22
Tx	ط	23
zx	ظ	24
3	ع	25
G	غ	26
F	ف	27
Q	ق	28
K	ک	29
G	گ	30
L	ل	31
M	م	32
n, N	ں،ن	33
o,v,w	و	34
H	ھ،ہ	35
’	ء	36
Y	ی	37
Y	ے	38

Urdu has no distinct upper and lower case letter forms [5]. However the Romanization scheme shown in Table 2 is case-sensitive (Roman letters only) that helps in distinguishing the correct Urdu pronunciation [4]. The table is arranged for reading from right-to-left in order to comply with the native Urdu reading style. Each Urdu letter is mentioned along its respective letter used for Romanization. Lower-case Roman letters represent the pronunciations exactly similar to their respective pronunciations in English. Upper-case letters represent similar but non-equal English pronunciation for the same letter.

3.3 Contextual shape changes of Urdu letters

Urdu letters change their shape based on their respective positions inside a word. A letter can have up to four different shapes i.e. base, initial, medial and final shapes.

Example:

A letter is in its base shape when it appears alone as a disjoint letter e.g. the letter “ج” pronounced as jim with IPA (International Phonetic Alphabet) “[d͡ʒ]” . Rest of the three shapes of “ج” are shown in Figure 3.

Fig 3 Contextual shape changes of letter “ج”

Initial shape refers to the shape of a letter when it appears in the beginning of a ligature. Medial shape of a letter is written when it is joined by both the preceding and the following letters inside ligature. Final shape appears when a letter marks the end of a word or ligature.

4. Design Parameters

At present, more and more data is being generated and uploaded using touch screen smart gadgets. These gadgets come in various shapes and screen sizes such as tablet PCs and mobile phones etc. Recently, there zero button touch screen laptop systems have been introduced in the market e.g., the Acer ICONIA. The current trends and types of new gadgets being introduced in the market suggest the growth of touch screen systems in the days to come.

Design constraints are not limited only to Urdu language and its specific features. There are some additional design issues also that are summarized in the following sub-sections.

4.1 Hygienic design

Full keyboard replica designs e.g. QWERTY and Dvorak etc with both base and shift versions cause usability problems as well as visibility problems hence not viable for touch screen systems particularly those with smaller screens. This problem becomes graver on smaller screens and large letter-set languages. The handheld touch screen devices offer very little screen area for keypad parking. This means that in QWERTY type keypads, the individual key size to type an Urdu letter becomes too small to clearly see and type with fingers. Thus such a keypad is more prone to errors during text entry. Besides, data input using small screen devices bring about health hazards to the user. Eyesight weakness, RSI (Repetitive Strain Injuries) and CTS (Carpal Tunnel Syndrome) etc. are only a few health hazards caused by technology/devices that we use in our daily life. For example, in case of eyesight, the closer objects put greater strain on the muscles converging the eyes retina [2]. Stress on convergence system of eyes is crucial factor for strain [15][25]. Thus we need to keep hygiene in prime focus during design and development of input systems, particularly for small touch screen devices.

We put forth hygiene in prime focus at the design time. Small devices put more strain on eyes due to acute and meager visibility [1][2][6][15]. RPA (Resting Point of Accommodation) and Convergence prospects were among important considerations at the design time. RPA deals with the point when the lens capsule changes shape to focus on a close object [15]. Convergence allows the image of the object(s) to be projected to the same relative place on each retina [2]. RSI (Repetitive Strain Injuries), CTS (Carpel Tunnel Syndrome), CTD (Cumulative Trauma Disorder) and ophthalmic endemics etc. are caused by regular and prolonged use of computers and its gadgets [25]. We developed distinct touch screen keypad that is "hygienic" to the users. At the same time, our design facilitates fast, correct and easy Urdu composing.

4.2 Virtual Keypads

Virtual keypad is also called soft keyboard [13][1]. Unlike the physical hardware keyboard(s), a virtual keypad shows up on the screen. Thus it consumes no physical space in the real world. However, it needs a much precious resource i.e. the screen area and uses some part of the same screen where data is typed i.e. the editor [1]. This gives rise to new concerns such as position, size, and orientation etc. of the virtual keypad w.r.t. the editor. We can make the virtual keypad context sensitive so that it is visible only when the user wants to input or edit text [13]. Theoretically we can show several distinct keypads at the same time, nonetheless a single user would use only one virtual keypad at a single time.

We borrow the assessment method of virtual keypads from the physical hardware keyboards evaluation technique. This comprise of two major parameters; i) the easiness to learn and ii) efficiency [13]. The former parameter takes into account the time needed for a novice to become a veteran with the keyboard whereas the latter parameter refers to the composing speed by a skilled user, a user well familiar with the system under study.

5. Contemporary and proposed keypads

Apart from the conventional QWERTY and Dvorak keyboards, there are a number of keypads used for text entry e.g. Multi-tap T9, odometer-like, touch-and-flick, Septambic keyer and Twiddler etc. Different keypads are suitable for different types and size of devices on which one or more of the above keypads are deployed [3]. It has been proved that the Dvorak keyboards are much better in performance than QWERTY keyboards [18]. However the generic QWERTY keyboards are still a major portion of the keyboards used around the world, perhaps due to the large amount of production and sales that already reached the consumer market.

5.1 Existing On-Screen Keyboard

A simple search query on any online search engine results in a large number of layouts for Urdu keyboards. However, there is no trace of research for such layouts. Therefore we considered the following two keyboard layouts for the comparison analysis of the existing systems and our proposed system.

5.1.1 Microsoft OSK (On-Screen Keyboard)

Microsoft Windows comes with a built-in soft keyboard called the OSK. It supports a number of languages including Urdu. It is basically a replica of the generic and classical QWERTY type hardware keyboard. This OSK is shown in Figures 4(a) and 4(b).

This OSK has migrated to many touch screen platforms including tablet PCs and smart phones. However, in our research we reached a conclusion that this keypad does not provide optimum performance and ease of use.

5.1.2 Phonetic Urdu Keyboard

The Center for research in Urdu language processing, Pakistan' introduced a phonetic keyboard layout for Urdu deployed on Microsoft Windows platform. The keyboard was originally designed for hardware keyboard layout with the on-screen version for typing guidance. The base and Shift-ON layouts of this keyboard are shown in Figures 5(a) and 5(b).

5.2 Proposed Keypad

Figure 6 shows our frequency based full keyboard layout for touch screen systems. The current layout for Urdu keyboards are replicas of the traditional QWERTY OSK (On-Screen Keyboards as shown in Figure 4(a) through Figure 5(b). This type of keyboard is not frequency based and has much room for improvement in that some high frequency letters are typed in combination of Shift-key, the last thing a user will need. Using Shift version of the keypads, shown in Figure 4(b) and Figure 5(b) would require double labor and double amount of time in typing.

In addition to the 58 letters shown in Figure 2, Ligatures and Diacritics are also borrowed from Arabic in Urdu. Ligatures are fixed blocks of letters each represented by a single Unicode. Each of them has been allocated separate key on the Phonetic keyboard as shown in Figure 5(a) and Figure 5(b). The unigram frequencies of Ligatures and Diacritics are very low. Therefore we allocated them a single button on our proposed keyboard layout. Diacritics are another set of low frequency characters. They are small macron-like characters normally used to show the correct pronunciation of letters in a word. Both the Ligatures and Diacritics are used mostly in religious texts that have become part of Urdu but they have been originally borrowed from Arabic and Persian.

We solved the aforementioned problems and proposed the frequency based keypad layout as shown in Figure 6. Based on the feedback from the volunteering subjects, the detailed performance examination such as homing time and response time to an external event of this keypad was not done. However the new layout has eliminated the Shift version of Microsoft Windows replica. We also re-arranged the position of keys based on the frequencies of individual letters such that the most frequent letters should be typed by the strongest typing finger i.e. the index finger. Another point in this arrangement is that the keypad is suitable for alternate hand typing which makes the typing much faster and easier. Using automated procedures to compare the existing and our proposed keypad would be a way forward.

There are some additional issues related only to the touch-screen keypads such as the inter-keys distance and the neighborhood of some standard reserved keys might also be required to change. One such example is the neighboring keys of the "Backspace/Delete" and the "Enter/Return" keys. During data entry, both these are opposite in function to each other as the former is used to delete a line whereas the latter is used for a line feed i.e. inserting a line. If user tries to touch-type, there are chances of touching a neighboring key by mistake. We suggest that "Backspace/Delete" and the "Enter/Return" keys should be designed and placed away from each other on touch screens. Also there should be some pixels left blank between every two neighboring buttons on the keypad to reduce typing mistakes. Our proposed keypad for large size touch screen devices is shown in Figure 4.5

Fig**. 4 (a) Base version of Microsoft Windows Vista OSK (On-Screen Keyboard).**

Fig. 4 (b) Shift version of Microsoft Windows Vista OSK.

Fig. 5 (a) Base version of Phonetic Urdu keyboard.

Fig. 5 (b) Shift version of Phonetic Urdu keyboard.

Figure 6 Proposed optimized Urdu QWERTY keypad. Shift version is eliminated.

6. Experiments

We carried out experiments on a general genre corpus of size 15,594,403 words. Using the unigram and bigram frequencies in a large corpus, we developed novel Urdu touch screen keypad as shown in Figures 3 and 4. The bigram characters neighborhood statistics reveal that the non-alphabetic frequency based arrangement of Urdu letters on our proposed keypad alone results in additional 9% improvement in the efficiency of our proposed keypad. The results of our experiments revealed ample significance that is explained in the comparison and evaluation Section i.e. Section 7.

6.1 Methodology

The methodology we adopted is enlisted stepwise in the following.

1. Calculate a frequency distribution for the words in an Urdu corpus of 15,594,403 words
2. Calculate a frequency distribution for the alphabets in the words i.e. the Unigram frequency distribution
3. Calculate a frequency distribution for the intra-words neighborhood of alphabets i.e. the characters bigram frequency distribution
4. Based on unigram frequencies, decide which alphabets will be typed by the strongest and weakest typing fingers
5. Based on bigram frequencies, decide the order of alphabets for alternate hand typing
6. Carefully design the input method keeping in mind certain additional factors such as health issues and Ergonomics
7. Compare the existing and proposed system using suitable statistical models

6.2 Model Used

In order to measure the efficiency of our proposed keypad, we use the model presented by Mark D. Dunlop and Finbarr Taylor [23].

$T(P) = T_h + w\,(K_w T_k + r(T_m+T_k))$

where

$T_h = 0.40s \rightarrow$ homing time

$T_k = 0.28s \rightarrow$ time required to press a key

$T_m = 1.35s \rightarrow$ response time to a word prediction event

$K_w = 5.421$ (U) $\rightarrow$ average length of an Urdu word (our modification in the original model)

w = No. of words

$r = 1.03 \rightarrow$ ranked word list selection time

To date, there is no full-fledged Urdu word prediction IME. In case of English and some other languages, existing touch screen systems start word prediction as soon as the user types the first letter. For words with length up to two letters, this seems to bring hardly any improvement to the typing speed. On the contrary, it makes the system more complex and larger in size putting more overhead on CPU. We recommend that word prediction should start after the second letter has been typed by the user. In the corpus we used, out of 15,594,403 words, 4,784,234 words are less than or equal to two letters in length. Hence for the experiments of this study, we discarded the words having length less than or equal to two character. The main reason to do so is; by the time the system is able to predict the desired word, the user will have already typed two letters or tapped the screen twice. Users evaluation showed that responding to a word prediction event and then tapping the appropriate option takes longer than typing the next alphabet from the keypad. Reducing the size of corpus gave us the extra advantage of using a smaller corpus of size 10,810,169 words that subsequently resulted in the low CPU overhead and less memory requirement for our proposed input system.

The bigram character neighborhood matrix of the entire corpus gifted us with an additional boost in typing speed in performance. Some Urdu words contain double and repeating letters. Using our proposed keypad the user needs to tap the same button twice in order to type a repeating letter. On the contrary, the same repeating letter can cost up to 12 taps in order to type it twice using a multi-tap T9 type of keypad.

7. Results, Comparison and User Evaluation

We compared the performance of proposed keypad with its existing counterparts. The evaluation was done by two distinct techniques; a) Automated performance evaluation b) Users evaluation.

7.1 Automated Performance Evaluation

The reduced corpus size and assumption of "touch=tap" put the bias in favor of the existing systems because a tap takes longer than a touch-and-flick. The same concept is applied to the base and the Shift-ON versions of the existing keypads. In our proposed keypad, due to elimination of redundancy and novel arrangement of individual letters on the keys, we eliminated the Shift-ON version and the user is still able to type the entire letter-set of Urdu. Thus we achieved results that show substantial improvement over the existing systems. The comparison of time required to type the corpus using existing Microsoft Windows and the Phonetic Urdu keyboards and our proposed keypads are illustrated in the Table 3. Thus the proposed keypad is 36.12% faster than its contemporary counterparts.

Table 3: Time analysis results chart

Time	*Existing Keypad*	*Proposed Keypad*
Seconds	539,930,226	344,886,062
Hours	149,980.62	95,801.68
Days	6,249.19	3,991.74
Improvement	**36.12%**	

7.2 Users Evaluation

Figure 7 shows the real world analysis through user evaluation.

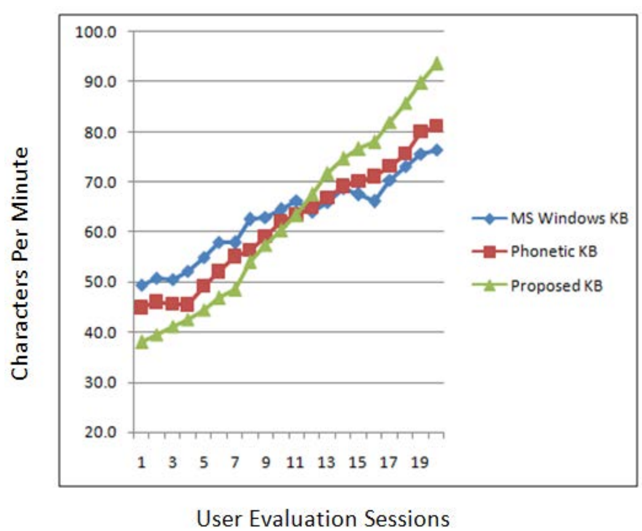

Fig. 7 Users evaluation results chart.

The user evaluation was carried out by three groups of native Urdu speakers. Two groups contained 15 males each and one groups comprised of 23 female subjects. All the subjects were college and university students and all of them volunteered their time and services for evaluation of this novel research. Their ages ranged from 18 to 26 years. 2/3 of the total subjects were right-handed and 1/3 left-handed. All of them were well versed with computers and experienced in typing but none of them was a professional typist. However, all of them had the experience of using the Microsoft Windows OSK for Urdu and the Phonetic Urdu keyboard. Most of them were familiar with the Multi-tap T9 Urdu mobile keypad also. This comparison will be included in our on-going research and will be published at some later time. The Acer Iconia running Microsoft Windows 7 was used as a test bed during user evaluation. Each user was allowed to re-size the width and/or height of the entire OSK keyboard to adjust the width and height of keyboard according to the size of his/her hands and fingers. Our proposed keypad was novel for all the three participants. Except for a 15-minutes initial briefing, no training sessions were conducted before the volunteers could use our proposed keypad for typing unseen Urdu text.

We conducted 20 typing sessions. A session means that each user was given unseen text to type on the Microsoft Windows OSK, the Phonetic Urdu keyboard and on our proposed keypad. The order to use the three keypads and the text to type by each user was all random. The text length was also kept random and the users were always given unseen text to type. This user evaluation procedure was adopted in order to prevent the bias in favor of a particular keypad and/or a user.

The results have been averaged and illustrated in Figure 8. X-axis represents the number of sessions while Y-axis means the typing speed of users in characters per minute. All the values in the chart are averages of all the three users who performed typing in a random order using random order of keypads and random pieces of text. As clear from the chart, the learning curve for our proposed keypad is the fastest to memorize. The users took only two sessions to learn it in order to surpass their speed of typing using a Multi-tap T9 keypad. This shows that our proposed keypad is easy to understand and memorize, hence user friendly.

Since the users were familiar with Microsoft Windows OSK and the Phonetic Urdu keyboard to type Urdu text, therefore the advantage was in favor of Microsoft OSK when we started user evaluation. Nonetheless, it took our novel keypad 9-user sessions to show better performance than the existing systems. During evaluation of our proposed keypad, the user evaluation did not show any significant difference between the working and performance of the diagonal and the horizontal neighboring letters on the keypad.

8. Conclusion

We proposed a novel keypad for small handheld touch screen devices. The comparison analysis were performed on two distinct tracks; the automated procedures and by detailed real life evaluation through user study. Both the evaluation method showed promising results. In addition to a significant amount of improvement over existing keypads, our proposed keypad design is flexible because the size and dimensions of keypads, buttons, and editors can be adjusted according to the device on which the keypad is to be deployed. Similarly our keypad offers greater usability because Urdu letters include all the letters of Arabic and Persian. Hence our keypad is equally usable by the Arabic and Persian users. The keypad is optimized for Urdu though. With minor additions, our input system is extendible to other Perso-Arabic languages as well. Similarly with small editing in the back-end script, our proposed system is equally applicable to non-Roman non-Perso Arabic script languages e.g. Hindi and Thai etc.

9. Future Directions

We intend to make our keypad public to research community for further extendibility to their respective native languages. More thorough testing of our keypad by a score of human subjects is also welcome. Additionally, we want to extend our keypad to include other Perso-Arabic languages such as Punjabi, Pashto, and Dari etc. Touch screen devices come in various shapes, screen sizes, hardware and software platform. We intend to develop optimized keypads for various touch screen gadgets such that each keypad best suits a certain type of gadgets. Our proposal of an optimized keypad for mid-size touch screen devices such as tablet PCs is already in its final stages of evaluation. Another possibility to exploit our work can be in the design of a single hand operated keypad (separate designs for each of the left and right hand), single finger operated and two fingers operated keypad designs suitable for numerous touch screen devices.

10. REFERENCES

[1] Andrew Sears, Julie A. Jacko, Josey Chu, and Francisco Moro, "The role of visual search in the design of effective soft keyboards", Behaviour and Information Technology, 2001, 20(3):159–166.

[2] Ankrum, D.R., "Viewing Distance at Computer Workstations", Workplace Ergonomics, September/October, 1996.

[3] Asad Habib, Masakazu Iwatate, Masayuki Asahara and Yuji Matsumoto, "Different Input Systems for Different Devices: Optimized Touch-Screen Keypad Designs for Urdu Scripts" in proceedings of Workshop on Text Input Methods WTIM2011, IJCNLP, 2011, Chiang Mai, Thailand.

[4] Asad Habib, Masayuki Asahara, Yuji Matsumoto and Kohei Ozaki, "JaPak IEOU: Japan-Pakistan`s Input English Output Urdu (A Case Sensitive Proposed Standard Input System for Perso-Arabic Script clients)", in proceedings of ICIET 2010, Karachi, Pakistan.

[5] Asad Habib, Masakazu Iwatate, Masayuki Asahara and Yuji Matsumoto. Keypad for large letter-set languages and small touch-screen devices (case study: Urdu). International Journal of Computer Science Issues, Vol. 9, Issue 3, No. 3, May 2012. ISSN: 1694-0814.

[6] Atencio, R., "Eyestrain: the number one complaint of computer users". Computers in Libraries, 1996, 16(8): 40-44.

[7] Colin P. Masica, The Indo-Aryan languages. Cambridge Language Surveys Cambridge: Cambridge University Press, 1993, Cambridge.

[8] George Weber, "The World's 10 most influential Languages", American Association of Teachers of French (ATTF), National Bulletin, 1999, vol. 24, 3:22-28.

[9] I. Scott MacKenzie, Shawn X. Zhang, and R. William Soukoreff, "Text entry using soft keyboards", Behaviour and Information Technology, 1999, 18(4): 235–244.

[10] M. Ijaz and S. Hussain. "Corpus Based Urdu Lexicon Development", in proc. the conference on Language & Technology (CLT07), Bara Gali Summer Campus, University of Peshawar, August 7-11, 2007, pp. 85-94.

[11] Jaschinski-Kruza, W., "Visual strain during VDU work: the effect of viewing distance and dark focus", Ergonomics 31, 1988, pp: 1449 – 1465.

[12] Jennifer Mankoff and Gregory D. Abowd, "Cirrin: A word-level unistroke keyboard for pen input", in Proceedings of the ACM Symposium on User Interface Software and Technology, 1998, pages 213–214.

[13] K. Knight and J. Graehl, "Machine Transliteration", Computational Linguistics, Volume 24 Issue 4, 1998, MIT Press Cambridge, MA, USA, pp: 599-612.

[14] Leonard J. West, "The Standard and Dvorak Keyboards Revisited: Direct Measures of Speed", 1998, Technical report, Santa Fe Institute, New Mexico, USA.

[15] Lewis M. Paul (ed.), "Ethnologue: Languages of the World", Sixteenth edition, 2009, Dallas, TX, USA.

[16] M. Afzal, S. Hussain "Urdu computing standards: Development of Urdu Zabta Takhti" INMIC 2001, LUMS Lahore.

[17] Malik, L. Besacier, C. Boitet and P. Bhattacharyya, "A hybrid Model for Urdu Hindi Tranliteration", in proceedings of Association for Computational Linguistics, International Joint Conference on Natural Language Processing (ACL-IJCNLP), 2009, Suntec, Singapore.

[18] Mark D. Dunlop and Finbarr Taylor, "Tactile Feedback for Predictive Text Entry", in proceedings of Conference on Human Factors in Computing Systems, 2009, Boston, MA, USA.

[19] M. Afzal, S. Hussain, "Urdu Computing Standards: Development of Urdu Zabta Takhti (UZT 1.01)", in proceedings of IEEE International Multi-topic Conference, 2001, Pakistan, pp: 216-222.

[20] NASA, NASA-STD-3000, "Man Systems Integration Standards", Volume I - Standards and Volume II, Revision B, 1995, National Aeronautics and Space Administration, Houston, USA.

[21] Nuray Aykin, Pia Honold Quaet-Faslem and Allen E. Milewski, "Cultural Ergonomics", Handbook of Human Factors and Ergonomics, Third Edition, 2006, pages 177–190.

[22] Robert W. Proctor and Kim-Phuong L. Vu, "Selection and Control of Action", Handbook of Human Factors and Ergonomics, Third Edition, Wiley Online Library. 2006, pages 89–110.

[23] S. Hussain, "Letter-to-Sound Conversion for Urdu Text-to-Speech System", 2004, Center for Research in Urdu Language Processing (CRULP), Lahore, Pakistan.

[24] T. Rahman, "Language Policy and Localization in Pakistan: Proposal for a Paradigmatic Shift", Crossing the Digital Divide, 2004, Khatmandu, Nepal.

[25] U.Afzal, N.I.,Rao, A.M.Sheri, "Adaptive Transliteration Based on Cross-Script Trie Generation: A case of Roman-Urdu", CLT-09 Lahore, Pakistan.

[26] Unicode. 1991-2001. Unicode Standard version. Online version: http://unicode.org/charts/PDF/U0600.pdf (Retrieved on Oct 30, 2012).

[27] Uta Hinrichs, Mark S. Hancock, M. Sheelagh T. Carpendale, and Christopher Collins, "Examination of text-entry methods for Tabletop displays", in Proceedings of the IEEE Workshop on Tabletop Displays, 2007, pages 105–112.

[28] Waldemar Karwowski, "The Discipline of Ergonomics and Human Factors", Chapter-1, Handbook of Human Factors and Ergonomics, Third Edition, 2006, pages 1-31