The Evolution of Number Entry: a case study of the telephone

Parisa Eslambolchilar[1], Julie Webster[1] and Gerrit Niezen[1]

[1] Computer Science Department, Swansea University, Swansea, UK

{p.eslambolchilar, j.webster,g.niezen}@swansea.ac.uk

Abstract. This paper details a user study to investigate serial digit entry on analogue and digital input platforms and the errors associated with them. We look specifically at the case of entering eleven-digit telephone numbers without a decimal point. The telephone is used as a platform for comparison, due to its clear evolution from a rotary dial to a pushbutton keypad and more recently, touch-based input. Thirty participants took part in a user study, which concluded that the touch interface was four times less accurate than the pushbutton and rotary dial interfaces. The latter two interfaces performed with similar accuracy; however, users were more than three times faster on the pushbutton keypad and recognized almost twice as many errors on the rotary dial. We have extended previous error taxonomies to include some errors relevant to sequences of numbers and built upon task-based guidelines found in the literature to suggest context-based design considerations.

Keywords: Number Entry, Interaction Design, Usability, Telephone Interfaces.

1 Introduction

The design of interactive systems has evolved over time in hopes of making them easier to use. Errors are typically blamed on a human factor, forgetting that the system should have been built to take them into account and attempt to minimise these types of problems [4]. The goal of interaction design is to aid the user in completing a task with a minimal number of errors or no error at all. An interaction that fails to accomplish this has likely failed at various levels.

Number entry is a task that is performed daily with little conscious thought. These number entry tasks seem trivial, but we still experience problems when performing them. Consider how many times we dial a phone number. It is rare to dial a wrong number, but it does happen occasionally and the results are potentially embarrassing. Similar problems occur with alarm systems where the wrong code results in a blaring sound or with personal identification numbers (PIN) for bank accounts that lock us out if entered incorrectly too many times in a row. All of these systems use a simple input method, such as a knob or keypad, laid out with either buttons or a touchscreen. In these cases, inputting a single wrong digit does not get us close to what we intended - it is instead entirely wrong. A wrong digit in a telephone number will result in

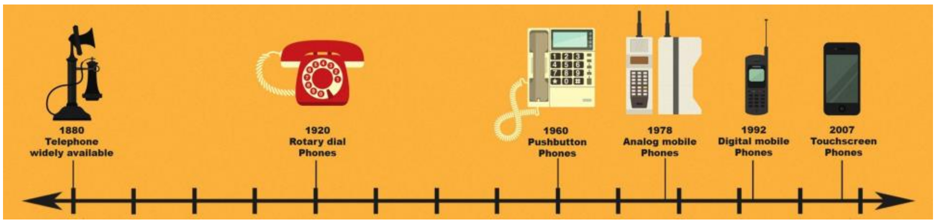

Fig. 1. The evolution of the telephone [5].

speaking to a different person all together. This specific category of numbers, which includes telephone numbers, PIN numbers, and credit card numbers do not have a decimal point. The telephone gives us a good vehicle for exploration because it demonstrates a clear evolution from one input device to another, as shown in Figure 1.

2 Related Literature

This section details literature that has influenced this case study of the telephone by looking at errors and their taxonomies as well as number entry interfaces.

2.1 What is Error?

James Reason [3] defines two major categories of error: slips (or lapses) and mistakes. *Slips* result from the incorrect execution or incorrect planning of a correct sequence of actions, so an error is made even though a person has the knowledge needed to perform the number entry task. In contrast, *mistakes* occur when a person has incorrect or absent knowledge of the task they are aiming to complete. In other words, mistakes result from the correct execution of an incorrect sequence of actions. For example, consider the misinterpretation of feedback: Reading “121” as the expected value “12.1” could be a slip, whereas thinking a symbol signifies “On” when it signifies “Start” is probably a mistake.

2.2 Taxonomies of error

Zhang et al. [7] proposed a cognitive taxonomy to categorise medical errors at the level of individuals, based upon Reason's taxonomy. Wiseman et al. [6] built on Zhang et al.'s taxonomy by focusing on number entry errors - errors that occur when a series of digits is being entered. In the medical domain, an example could be specifying values for an infusion pump, such as rate of infusion or the total volume to be infused. The study in [6] consisted of 20 participants each entering 30 numbers, where the length of numbers ranged from two to five digits. Half of the numbers entered were integers and the other half contained decimal points. A total of 350 errors were gathered and categorised into 21 classes of error, based on Zhang et al.'s taxonomy.

2.3 Number Entry Interfaces

Oladimeji [2] identified three categories of number entry interfaces:

- Serial digit entry – number is entered sequentially from left to right, e.g. telephones, ATMs or calculators,

- Independent digit entry – user controls the digits that make up the number separately, e.g. using up/down buttons to change the digit (0-9) and left/right buttons to change the position of the digit being entered, and
- Incremental number entry – user increases or decreases a number analogous to scrolling through values on a number-line, e.g. using up/down buttons, sliders or knobs.

There are certain trade-offs to consider when selecting a number entry interface, including speed of entry, severity of errors and space required for the interface. Oladimeji et al. [1] compared keypad interfaces to up/down button interfaces (press and hold to increase or decrease values) and showed a factor of two fewer unnoticed errors with up/down buttons. Fixation experiments showed that users are less likely to notice errors on the display when using keypads, as they fixate twice as much on the keypad itself (compared to the up/down buttons).

We focus our research on serial digit entry and the errors associated with these types of interfaces, when not using a decimal point. We focus on three types of serial digit number entry interfaces: the rotary dial, the pushbutton keypad and the touch display. The telephone was chosen as a platform for comparison due to its clear evolution from analogue interfaces to digital interfaces.

3 Experimental Setup

Existing studies on number entry research focus on a specific domain, for example entries on infusion pumps in the medical domain [2, 6]. The focus of this study was on more everyday interaction - serial digit number entry without a decimal point. This is the type of number entry we perform almost on a daily basis, for example entering PIN numbers or telephone numbers, where one wrong digit invalidates the whole number being entered.

A user study was performed to compare a rotary dial interface, touch interface and a keypad interface (Figure 2). A total of thirty participants completed the user study task. All were recruited through emails sent to the students and staff at the university and were affiliated with the university in some way. Most were postgraduate students and staff, 14 were male and 16 were female. Three participants were between 17 and 20 years old, 19 participants where in the 21-30 age group, four participants were between 31 and 40 years old, and four participants were between 41 and 50 years old. Participants were fluent in English for questionnaire purposes and were compensated for their time with a £5 shopping voucher. Before the sessions, the order of interfaces was randomly assigned and 50 11-digit numbers were randomly generated to allow all 30 users to enter the same 50 numbers on all three interfaces.

Participants were allowed five practice runs to familiarise themselves with the interface before the 50 11-digit numbers that were then recorded. The time it took to enter each number was recorded, along with the number entered compared to the number given to the participant. Participants were not provided with a display to see what digits they had entered, so they had to rely on the inherent tactile and visual feedback provided by the input devices. This was done to maintain consistency betw-

Fig. 2. The three telephone-style interfaces used by participants: the rotary dial (left), pushbutton keypad (middle) and touchscreen keypad (right).

-teen the three interfaces, as a visual display would not have been used on older interfaces.

Participants were not allowed to correct their entries. If they perceived that they made an error, they were asked to notify the researcher and continue entering the rest of the number. This made it possible to determine how many errors would have been recognised, for a given number being entered. To alleviate fatigue, participants received a five-minute break before using the next interface. After thirty user sessions, data was collected from a total of 4,293 numbers, amounting to 46,807 digits in total, including digit additions and omissions due to errors. These results were analyzed in a variety of ways, detailed in Section 4.

3.1 Hardware Setup

A rotary dial interface from an early 1960s telephone was used as the analogue input device. When the dial spins, a cog inside moves the same number of clicks as the number selected, opening and closing a switch. For example, moving the number "6" around and letting go will cause the cog to engage a switch six times. An Arduino Uno microcontroller was used to count the number of pulses and each digit was outputted to the computer along with a timestamp.

A button interface for a FEZ Spider Kit[1] was used as the digital input device. The interface consisted of membrane buttons laid out as a traditional 3x4 keypad and was connected to the kit. An SD card module was used to store the numbers entered on the device, along with a timestamp for each digit.

The touchscreen interface was created on a resistive touchscreen with the FEZ Spider Kit. A 3x4 keypad was created on the touchscreen to look like the physical keypad. The LCD, touch-capable screen was attached to the mainboard and USB-powered modules of the kit. An SD card reader was utilized in the same way as on the pushbutton interface.

[1] The FEZ Spider Kit is a .NET Gadgeteer electronic set produced by GHI Electronics, consisting of a mainboard and a number of attachable modules.

Table 1. Total incorrect digits and correlation of error rate and order of interfaces

Interface	Total Incorrect Digits	Correlation
Rotary	190 (1.19%)	0.344
Button	180 (1.12%)	0.615
Touch	896 (5.81%)	**0.046**

A small button labelled "Next" was placed next to all three interfaces, which the user pressed to indicate that he/she was ready to enter the next number. The interfaces were mounted on platforms (Figure 2) to allow users to interact with the devices in the manner that they were the most comfortable.

4 Results

A digit-by-digit analysis was conducted of each number entered to determine what types of errors were made. Table 1 summarizes the total number of digits that were incorrect in each interface.

We used a subset of the error taxonomy of Wiseman et al. [10] to classify the errors made in our user study. Wiseman's taxonomy included numbers with decimal points, so some categories did not apply to our study. The following definitions were used:

- Digit(s) wrong – a different digit was entered from what was intended (e.g. '1' instead of '4')
- Digit missing – a single digit was omitted (e.g. '14' instead of '124')
- Digit added – a single digit was added (e.g. '124' instead of '14')
- Anagram – digits were reversed (e.g. '14' instead of '41')

Based on observations during in our study, we have added the following error classes to more fully understand user errors when entering whole numbers:

- Repeat digit missing – a single digit the same as the previous one was omitted (e.g. '14' instead of '144')
- Repeat digit added – a single digit the same as the previous one was added (e.g. '144' instead of '14')
- Repeat *n* digit pattern – a pattern *n* digits long was repeated (e.g. '1414' instead of '14')

Table 2 shows the total number of each type of error on each interface, as well as the frequency of errors on the interfaces. In the tables, n represents the number of errors made and r is the frequency, calculated to be $100n/N$ percent to represent the frequency that particular error occurred on that interface.

The most common error on the rotary dial interface was entering an incorrect digit. Missing digits were the next most common error, but they were almost eight times less likely than entering the wrong digit. A total of 180 digits were entered incorrectly on the button interface, slightly less than the rotary dial and more than five times less than the touch interface. The most common error on the button interface was adding one repeat digit. This may have occurred because users held down a button for too long hoping to make sure that it registered, but instead resulted in multiple instancesof the same digit. A missing digit was the next most common error. The most comm-

Table 2. Rotary Errors (R), Button Errors (B), Touch Errors (T)

Error Type	nR	rR (N=15928)	nB	rB (N=15906)	nT	rT (N=15389)
Wrong digit	120	0.75	33	0.21	150	0.97
Digit missing	28	0.16	34	0.21	509	3.31
Added 1 repeat digit	19	0.12	94	0.59	30	0.19
Added 1 digit	13	0.08	8	0.05	87	0.57
Added 2 repeat digits	4	0.03	-	-	-	-
Anagram	3	0.02	3	0.02	19	0.12
Repeat digit missing	-	-	6	0.04	96	0.62
2 digit pattern missing	2	0.01	-	-	4	0.03
Repeated 2 digit pattern	-	-	1	0.01	-	-
Repeated 3 digit pattern	1	0.01	1	0.01	1	0.01

-on error on the touchscreen was a missing digit, likely a result of the user not tapping the screen hard enough to register their selection. A wrong digit and missing a repeat digit were the next most common errors but were roughly four and five times less common, respectively, than the omission of a single digit.

We conducted a mixed model analysis for each interface to determine if learning or fatigue affected the user as they completed the number entry task (Table 1). The analysis was performed by comparing the order that the interface was given to that participant and the error rate. Performance when the participant used the touchscreen first had a significance of 0.046 compared to a value of 0.022 when used second. No significant difference was observed for performance on two other interfaces.

Figure 4 shows a binary representation of all of the errors made on each interface, where each row represents a participant and each column represents a number being entered. Effects of learning would suggest that errors were concentrated at the beginning, while effects of fatigue would suggest errors concentrated at the end. Based on visual analysis of the figure, it appears that the rotary interface may have caused fatigue because there was a slightly higher concentration of errors at the end. In contrast, the touchscreen interface looks like it may have benefitted from learning as the user progressed through the task because the errors occurred slightly more often at the beginning. The button interface does not appear to have been affected by either because the errors look evenly spread across the use of the device.

5 Discussion

There are trade-offs to consider when selecting a number entry input device, as demonstrated by the data collected in our experiment. Our results showed that the button interface is only 0.07% more accurate than the rotary dial. Because the error rate is very close between these two interfaces, other factors such as the recognized error rate or the speed of entry may become important factors in selecting the most

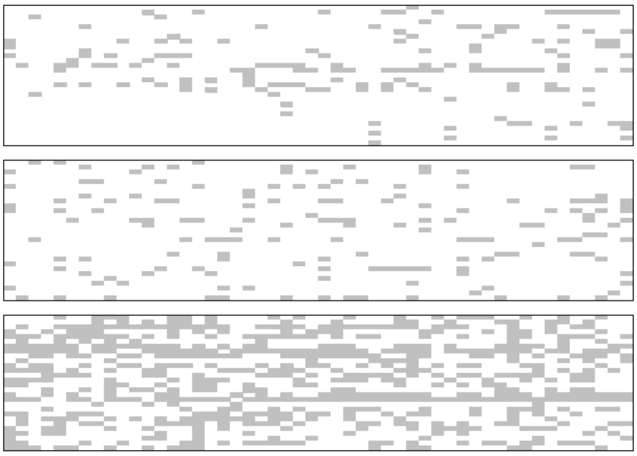

Fig. 4. A binary representation of all errors made on each interface: rotary (top), button (middle) and touch (bottom). The rows represent the thirty participants and columns are 50 numbers.

appropriate input device. Based on our results, the button interface is slightly more accurate may not always make it the right choice, because the recognized error rate was 13.80% higher on the rotary dial.

The most common errors on each interface vary and show that the errors must be dependent on the interface and therefore, independent of the user. This kind of analysis is important because we can note that some errors only occur on certain interfaces, or some occur on all interfaces at the same rate, such as ‘Anagram’, the reversal of two digits. This means that although it may not be possible to select an input device to reduce ‘Anagram’ errors, selecting the rotary dial will likely reduce ‘Repeat digit missing’ errors since they did not occur on the interface when tested in our study. By knowing what kinds of errors are common on each interface, it is possible to design devices that should work best for the objectives of that task.

One important observation from our study came from users’ interaction with the resistive touchscreen. Most touchscreen mobile phones and tablets on the market today use capacitive screens, which only require a tap of the fingertip to register the selection. Our resistive screen is more similar to the types of touchscreens used on other types of devices, such as ATMs, PIN pads and GPSs. The error rate in our study may have been so high on the touchscreen because users simply tapped the screen expecting the digit to register and did not watch for the visual feedback provided by the button on the screen to confirm the correct number. This would explain why the most common error was ‘Digit missing’, because users did not tap the screen hard enough, even if they thought they had entered the number correctly.

6 Conclusions and Future Work

In this paper we focused on analysing the number of errors that occurred during interaction with three number entry interfaces on a telephone platform. Telephone inputs are not outdated; on the contrary we use numeric keypads everyday at cashpoints.

Although physical rotary dialers are nearly gone, they are available as smartphone apps.

From our study, it is evident that multiple trade-offs need to be considered when selecting an input device, with the type of task not being the only design criterion. There are a number of contexts where the most appropriate design for the user may be different than what is best for the task, so a compromise must be made. Instead of compromising to sacrifice usability for one group, considering the context could instead create a successful design. Many users have memorized a certain position of their fingers on a PIN pad to quickly enter their code, so if the layout was changed from the usual '1' at the top to having the '1' at the bottom, many users would enter their PIN numbers incorrectly in the context of paying at the till in a busy shop filled with distractions.

By considering learning and fatigue as factors in a device's use, an informed decision can be made about the most appropriate input device for the task. Since one device may have shown learning, one fatigue and one neither, we assume that the interfaces themselves caused the errors. This suggests that the interface choice is critical to the usability and performance of a device.

As future work we would like to repeat this study using other platforms of number entry, such as capacitive touchscreens, to find out if the touchscreen error rates are still so much higher than the other two interfaces. We would also like to compare these results with the errors caused by continuous input method of entering numbers such as knobs. This would lead to providing generalised recommendations for number entry interfaces. Devices could also be fine-tuned to offer more intuitive interaction for users. For example, incorporating real-world physics into an input device's response could assist users, such as a rotating wheel on a touchscreen that mimics the effects of friction when slowing down.

References

1. Oladimeji, P., Thimbleby, H., Cox, A.: Number Entry Interfaces and their Effects on Error Detection. In: Proceedings of the 13th IFIT TC 13 International Conference on Human-Computer Interaction, vol. 4, pp. 178--185 (2011)
2. Oladimeji, P. Towards Safer Number Entry in Interactive Medical Devices. In Proceedings of the 4th ACM SIGCHI symposium on Engineering Interactive Computing Systems (EICS'12), pp. 329–332 (2012)
3. Reason, J. Human error. Cambridge, UK: Cambridge University Press, 1992.
4. Thimbleby, H.: Press On: Principles of Interaction Programming. The MIT Press, Cambridge, Massachusetts (2007)
5. Webster, J, Eslambolchilar,P., Thimbleby, H.,. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pp. 596--597. ACM Press, New York, New York (2012)
6. Wiseman, S. Cairns, P. Cox, A.: A Taxonomy of Number Entry Error. In: Proceedings of the 25th BCS Conference on Human-Computer Interaction, pp. 287-196 (2011)
7. Zhang, J, Patel, V.L., Johnson, T.R., Shortliffe, E.H.: A Cognitive Taxonomy of Medical Errors. In: Journal of Biomedical Informatics, vol. 37, pp. 193-204 (2004)