



HAL
open science

Increasing Accuracy by Decreasing Presentation Quality in Transcription Tasks

Frank Soboczinski, Paul Cairns, Anna L. Cox

► **To cite this version:**

Frank Soboczinski, Paul Cairns, Anna L. Cox. Increasing Accuracy by Decreasing Presentation Quality in Transcription Tasks. 14th International Conference on Human-Computer Interaction (INTERACT), Sep 2013, Cape Town, South Africa. pp.380-394, 10.1007/978-3-642-40480-1_25 . hal-01501757

HAL Id: hal-01501757

<https://inria.hal.science/hal-01501757v1>

Submitted on 4 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Increasing accuracy by decreasing presentation quality in transcription tasks

Frank Soboczenski¹, Paul Cairns¹ and Anna Cox²

¹ University of York, Department of Computer Science, Deramore Lane,
YO10 5GH, York, UK {fs596, Paul.Cairns}@york.ac.uk

² University College London, Interaction Centre, Gowners Street,
WC1E 6BT, London, UK Anna.Cox@ucl.ac.uk

Abstract. Many tasks with interfaces require entering data accurately, for example, entering patient data in an electronic records system or programming infusion pumps. However, human error is inevitable. Paradoxically, results from cognitive psychology suggest that representing information in a poorer quality format increases the likelihood of memorising the information accurately. This is explained by the dual system account of cognition where slower, more effortful but more accurate thinking is invoked via the poorer quality representation. We present two studies where we transfer these results to the domain of data-entry and show that poorer quality format of to-be-copied information leads to increased accuracy in transcription tasks. Moreover, this is not a consequence of the typical speed-accuracy tradeoffs. The results of our novel approach have implications for the design of data-entry tasks in domains such as healthcare.

Keywords: Human error, cognition, data-entry, presentation quality, number-entry, infusion pump, perception, safety-critical systems

1 Introduction

Routine data transcription tasks occur in a large number of settings, e.g. entering student marks into a spreadsheet, and paying bills via an online-banking system. But people are also prone to errors and while a wrong letter or number is sometimes little more than an inconvenience, things can get pretty ugly in safety-critical domains. One such domain, depending on a high standard of accuracy and data quality, is healthcare. For example, nurses are often required to enter large amounts of sensitive data in electronic health records (EHRs) and to program drug infusion pumps with information located on a prescription or infusion bag.

A simple slip while transcribing patient data or programming a syringe pump in a hospital can quickly turn a routine task into a serious situation. Previous studies have shown that human errors are the source of more than 50% of incidents in most domains and in aviation can be as high as 90% [20]. In healthcare the process of studying errors is only now emerging [19]. Recent studies have shown that people make transcription errors frequently and many systems do little to detect and prevent errors

[18]. These issues strongly motivate our interest in reducing error rates in transcription tasks and we view this work as demonstrating an important new approach in this area.

Transcription tasks require users to perceive the target information, which is then coded into easily remembered parts in their memory. This is followed by the decomposition of the remembered chunks into separate characters to associate the matching keystroke to the relevant character. Once this association is complete, the individual characters are then translated into movements to press the proper key during the data-entry process [14]. Previous research has shown that the ease with which a user can access the to-be-copied text impacts the strategies used to complete the tasks. So, how far away the to-be-copied content [3] is or whether or not the user has to click to reveal it [8], can lead to differences in accuracy of performance. This is explained by the Soft Constraints Hypothesis [8], which suggests that people are more likely to use a potentially less accurate *memory-intensive* strategy over a more accurate *perceptual-motor* one when the cost of accessing information is higher. Gray et al. (ibid) demonstrate that memory-intensive strategies can lead to higher error rates in transcription tasks and therefore argue that reducing the effort required to access information can be beneficial.

Whilst much research has focused on decreasing the costs of accessing information, recent research has demonstrated that there may be disadvantages to this approach if the information processing strategy adopted is too passive [11]. Supporting evidence from [5] suggests that *increasing* the effort required to access information to be transcribed can be beneficial. Specifically, they argue that information represented in a less readable format can increase the likelihood of accurately memorising the information [5]. One explanation for this finding is that, in order to make sense of the text or the numbers, the user experiences cognitive dysfluency when reading and so moves away from more automatic reading processes towards more effortful processes [2]. Different names are used to describe these kinds of processes such as for example *implicit* and *explicit* [7, 13] and the more neutral, and increasingly common terms, *System 1* vs. *System 2* [17]. *System 1* refers to automatic, hard to control, effortless and fast processes [10]. In contrast, *System 2* is slow in execution, effortful and is limited in its capabilities [10]. Whether a *System 1* or a *System 2* process is chosen to complete a task is influenced by the task at hand. For example, when a calculation task changes from a relatively simple to a difficult task *System 2* takes over. An example of such a scenario is thinking of tasks such as the question of what is the result of 2×2 or the question of what is the result of 342×319 . In the latter example, *System 2* takes control over the thinking process to produce an answer. The literature also presents an expansion of these well established thought processes by a third one namely *System 3*. This third process reflects on complex problem solving and prolonged major decisions, which take a very long time (i.e. months or years) and are therefore not applicable in this experimental context [6].

There are also attempts in the literature to examine ways to stimulate *increased cognitive engagement* in certain situations. In one particular study [5], researchers tried to provoke people to think harder by manipulating the ease with which fonts could be

read, and measuring the impact on the participants' ability to memorise specific information. Their results indicate that there is potential for great improvement in performance of people's memory when presenting information in a harder to read font.

In this paper, we present two studies to explore the effects of *System 1* vs. *System 2* on data transcription tasks. This is a novel application of results of [5], which are about memory, to the domain of interaction by which we might contribute to the findings of Gray et al [8]. In our first study we investigate the effects of dysfluency of text in the domain of text-entry. We chose text-entry as an initial study because it is widely known that number-entry studies produce a very low error presentation rate (usually below 5%) [12,18]. As we are specifically looking for errors, a text-transcription task seemed ideal to simply get enough data to work with. The research instrument in the first experiment was the text-entry game *TypoMadness*, which we designed specifically for this purpose (see Fig. 2). Participants had to transcribe text phrases as quickly and as accurately as possible in two different conditions *Clear* & *Obscured*. In the second study, we transferred the results from text- directly to number-entry, comparable to the tasks in a healthcare environment. The task for participants in the second study was to re-enter given numbers in a calculator style interface i.e. serial interface in three different conditions *Clear*, *Medium* and *Obscured*.

2 Text-Entry experiment

At first pass, it appears intuitive that text that is less easy to read (illegible or dysfluent text) would result in a higher number of errors in a text-transcription task as a user may be more likely to misperceive the text. However, evidence from [5] & [11] suggests that dysfluent text could encourage the adoption of an alternative information processing strategy, resulting in fewer errors. The hypothesis for this study was that people would make fewer errors and therefore archive a higher score in the *Obscured* condition than in the *Clear* condition.

2.1 Participants

20 participants (9 men) with a mean age of 27.2 years (st.dev = 8.5) and mostly students at the University of Anonymus were randomly recruited by email as well as personal invitation. Nine participants stated that the task was difficult. However, only four of these nine people were part of the *Obscured* group. Moreover, these four participants further stated other reasons such as being unable to use the backspace-key than the visibility of the given text as the reason why they felt the task was difficult. None of the participants had dyslexia or any other disability related to vision and reading. Additionally, all of the participants stated that they had normal or corrected to normal vision.

2.2 Design

A between-subjects design was used with ten participants in each group. The independent variable is the grade of visibility of the text. There were two conditions: *Clear*, which used clear visible content, (font-colour; #000000) and *Obscured*, which used less clear i.e. less visible content (font-colour: #BABABA) (see Fig. 1).

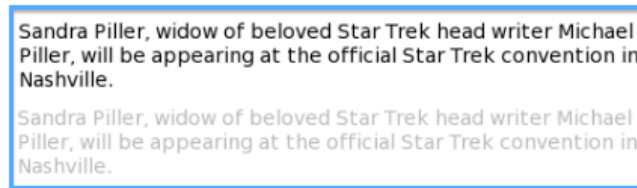


Fig. 1. Example of the text displayed in the *Clear* (upper) and *Obscured* (lower) conditions.

The dependent variable was the score. It is the accumulated sum of correct entered characters in each task reduced by errors. An error in this study is anything that is not identical to the target text. This could be an additional space, upper or lower case difference, wrong or missing letters generally any sign different from the displayed text. Due to limitations of space, we have not reported an analysis by error type (slip, mistake etc.) in this paper. The set of 26 text phrases was counterbalanced across the study to avoid ordering effects.

2.3 Materials

The task was framed as a game in which the aim was to enter as many sentences as accurately and as quickly as possible within 15 minutes. Each sentence had between 95 and 161 characters and was displayed in either *Clear* or *Obscured* format. Additionally, the sentences were not connected in prose or any other form, so participants could not have predicted them. The sentences were randomly chosen from news and boulevard press websites such as theguardian.co.uk or cnn.com.

The game used for this study was implemented as a website in PHP with JavaScript. The website was connected to a MySQL database allowing the entered text phrases to be stored in the database for each participant. The displayed score was calculated out of the matching characters of each submitted text phrase. Each score for individual text phrases was stored in the database and displayed via cookies on the game website. The game website consisted of a display area in the centre of the page where the text paragraphs were displayed, a textbox for typing the text and a red button to submit the entered text and continue with the next task. The backspace button and the cursor keys were disabled in the text field so that participant could not change their entered text. All entered sentences no matter what they might be were saved in the database after a participant pressed the red button. The entered sentences were saved as a string in the database and compared to the original text. If there was any difference the score for that task was modified accordingly. It should be noted that

where an error arose due to inserting an extra character, this did not cause all subsequent characters to be classed as erroneous.

The game was played on a MacBook Pro with a UK English keyboard layout and a 13" size display at 1280x800 resolution. The browser used for the experiment was Google Chrome running under Apple Mac OSX 10.7.5 (Lion). The game website ran in the browser, the browser window was maximised and controls were disabled from being shown so that only the game itself was visible. All other applications were closed. In addition, there was also an informed consent to sign at the beginning of the study and a brief paper-based demographic questionnaire to fill out by participants after they completed the main task. The demographic questionnaire was used to specify the participants but not further used in the analysis reported here. A smartphone was used to stop the time by the experimenter.

2.4 Procedure

After reading the instructions, participants completed two training sentences, which were separate from the main sentence set of the experiment. The target text was presented in a textbox in the centre of the individual webpage. After the participants finished entering the training session a new start page was displayed only with a start button. Participants could see their score (in form of a character count as seen in Fig. 2) displayed on the right side of the text field. On completion of the training, the participant proceeded to the main study by clicking on the start button.

The target text was presented either *Clear* or *Obscured* depending on the condition. Participants were instructed to enter the target text as quickly and as accurately as possible. In order to get the highest possible score, all tasks had to be completed within 15 minutes though participants were not shown how long they had been playing. The experimental design in form of a game made it necessary to put participants under some form of time-pressure. This was also done as means of increasing the error rate. Each time the participant pressed the red button, their score was updated though there was no direct indication of any errors. If they completed the task within the 15 minutes, the final pages displayed the overall score. The experimenter controlled the smartphone and stopped the participant at the end of the given time limit. After finishing the entry task, the experimenter offered each participant a small break before asking the participants to fill out a demographic questionnaire. The participants were then debriefed as to the goals of the study.

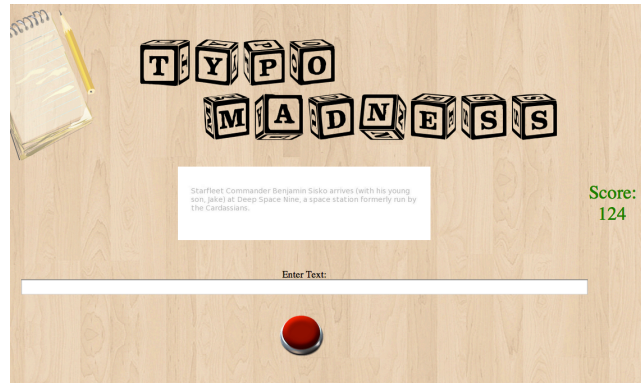


Fig. 2. The interface of the game *TypoMadness*.

2.5 Results

The hypothesis for this study was that there would be significantly fewer errors in the *Obscured* group (i.e. a higher score) than in the *Clear* group. The mean of the errors made in the *Clear* group was 830.1 (sd = 516.3) whereas in the *Obscured* group was 360.4 (sd = 127.0). The spread of errors is much higher in the *Clear* group than in the *Obscured* group (see Fig. 3).

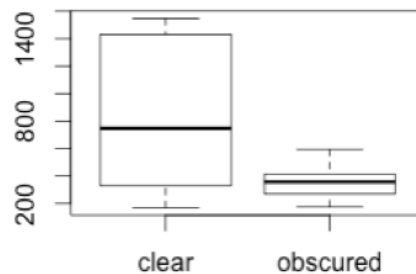


Fig. 3. Boxplot of total number of errors made by all participants in each condition.

A Mann-Whitney-W test was conducted, as error data does not follow a normal distribution, which revealed a significant difference between the two groups ($W = 23$, $p = 0.04$). Furthermore, though we had a directional hypothesis, a more acceptable two-tailed test was chosen to avoid the controversies around one-tailed testing [1]. This move also resulted in a higher threshold for significance and therefore would give us more confidence in our findings.

In terms of speed of completing the task (see Fig. 4), the mean rate for entering sentences for the *Clear* group was 1.69 sentences/min (sd = 0.45) and for the *Obscured* condition 2.02 sentences/min (sd = 0.49). There was no significant difference in the rate of entering sentences ($W = 26$, $p = 0.07$).

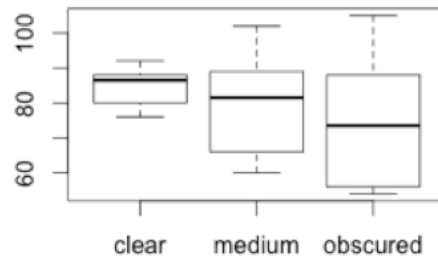


Fig. 4. Boxplot of the rate of typing (sentences/min) of participants in each condition.

Moreover, the participants in the *Clear* group entered a mean of 23.1 (sd = 2.99) sentences and the in the *Obscured* group 25.0 (sd = 1.63). A boxplot illustrates the sentences entered by group (see Fig. 5). There were also no significant differences between the number of sentences participants completed ($W = 32$, $p = 0.15$).

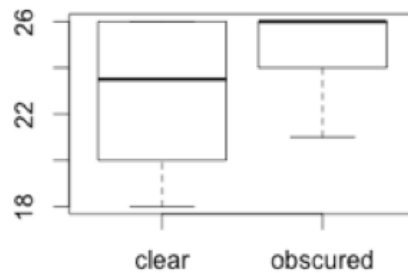


Fig. 5. Boxplot of the number of sentences entered by each participant in each condition.

Four participants from the *Clear* group and seven participants from the *Obscured* group managed to enter all the 26 given text phrases and finished ahead of time. This leaves six participants from the *Clear* and three from the *Obscured* group who could not finish in time and therefore used the full 15 minutes.

2.6 Discussion

As predicted, the results showed that participants in the *Obscured* group achieved higher accuracy scores (i.e. fewer errors) than those in the *Clear* group. One explanation for this reduction in errors could be a speed accuracy trade-off. However there was no significant difference between the two conditions in neither the rate of enter-

ing sentences or on the total number of sentences entered by each group. Indeed, the participants in the *Obscured* group entered more sentences, working at a faster rate suggesting that the reduced readability of the font was not slowing the participants down. Nonetheless, they made fewer errors than the *Clear* group.

Our manipulation is identical to that of [5] who suggest that *increasing* the effort required to access information results in an increase in the likelihood of accurately memorising the information [5]. However, we only focussed on representing the to-be-entered information in a faded font-colour but not also in a different font-style, which indicates a more robust outcome. The results therefore suggest that a more memory intensive strategy may have been adopted by the participants in the *Obscured* group which, in this case, resulted in enhanced performance.

Focusing on limitations, it can be argued that the size of our sample (20 participants) is simply too small to have confidence in our data. This might be true and requires further studies to get a more robust set of data, which is one reason why we decided to present a second study on the effect of *System 1* and *System 2* in this paper. On the other hand a larger sample size might just present a larger effect of our detected result as we have chosen to examine our data with a higher threshold of significance (two-tailed test). An improved experimental design would suggest to use sentences, which are all of the same length to avoid possible control confounds. Additionally, grey text can be more legible than black text under specific circumstances, depending on the level of contrast with its background and the overall level of luminance. Some types of dyslexia respond well to contrast reduction for improving readability. However, there were no dyslexic participants (at least not that they were aware of) in our study. Additionally, prior experience in data-entry work of participants or even any prior knowledge of the material might have influenced their performance in our task; therefore a more careful selection of participants for future studies needs to be considered. Moreover, there is also the argument that a too high grade of obscurity of text may result in participants getting frustrated or suffer from fatigue and therefore abandon the task. We regard this as major issue and designed our next study to explore performance across a range of levels of obscured information.

3 Number-Entry experiment

As the results from our initial text-entry study were very promising we were keen on exploring the discovered effect of *System 1* vs. *System 2* in number-entry. Therefore, we directly transferred the task from text- to number-entry in the next experiment. The hypothesis for our second experiment was that there is a significant difference in the number of errors between the conditions *Clear*, *Medium* and *Obscured*.

3.1 Pilot

The reason why the number-entry study consists of three conditions resulted out of an initial pilot study among 53 first-year undergraduate students in computer science. This pilot to measure the effect of *System 1* vs. *System 2* in the domain of number-entry was designed accordingly to the previous text-entry experiment with two conditions *Clear* and *Obscured* to measure a significant difference in the amount of errors between these two groups. We discovered that participants did make more errors in the *Clear* condition but not significantly more errors. We also noted that our initial design of our *Obscured* condition was too strong i.e. the obscured numbers were too faded, which made it too hard to read for participants. It was then decided to design this study with three conditions to explore a wider range of different levels of obscurity. These three conditions were all less obscure than in the pilot study. The participants of this pilot study were not allowed to take part in the described follow-up study.

3.2 Participants

Participants were randomly recruited by personal invitation or by email. There were 30 participants (16 men) with a mean age of 26.5 years (st.dev = 8.2) in total. Participants were mostly familiar with a calculator based number-entry interface as used in the study but none of them had previous experience of working in the healthcare sector. However, three participants stated that they were familiar with infusion pumps. Two people stated that they found the task difficult (one in the *Clear* and one in the *Obscured* group) and two different participants stated that they did not consider themselves to have normal vision (also one in the *Clear* and one in the *Obscured* group). Five participants had never used a number pad before, 7 rarely used it, 14 sometimes used it and only 4 participants (one each in group *Clear* and *Obscured* and two in group *Medium*) often use a number pad to enter numbers.

3.3 Design

A between-subjects (3 groups) design was used with 10 participants in each group. The independent variable is the grade of visibility of the presented numbers and the textbox where participants enter the numbers. There were three conditions: *Clear*, which used clear visible content (font-colour; #000000), *Medium* (font-colour; #BDBDBD) and *Obscured* (font-colour: #DFDFDF), which used less clear i.e. less visible (light-grey colour) content (see Fig. 7). The dependent variables were the total number of errors made, particularly the rate of errors and also the number of corrected errors in each group. A single error is counted when any number is entered other than the displayed number. Further, a corrected error is an error where participants noticed that they made an error and entered the number again. These errors were not added to the total number of errors. As previously, error type is not considered in our analysis.

3.4 Materials

The task set for participants was designed as a game in which the aim was to enter as many numbers as accurately and quickly possible in a calculator-based interface within 5 minutes. Compared to the first experiment a five-minute limit was chosen for this study as the pilot study showed that participants could transcribe content that consists of numbers considerably faster than text. The visual design of the serial i.e. calculator-based number-entry interface was carefully created after real world infusion pumps such as the Baxter[®]

AS40A or the Graseby[™] 3400 syringe pump (see Fig. 6). For the purpose of this study only the number pad of the infusion pumps with an *Enter* and a *Clear* key was replicated all other buttons of the original interface design were removed from the interface as they could have a possible unwanted effect on participants.



Fig. 6. Infusion pump models Baxter[®] (left) and Graseby[™] (right).

The interface was implemented in PHP and JavaScript with a logging feature, which recorded all user interactions with the virtual device and wrote all interactions into a MySQL database. The interface provided thirteen functional buttons usable by the keyboard-only interaction (numpad).

The display itself consisted of 16 components. Three number displays (number to be entered, score and number input field), *Enter* and *Cancel* buttons and 0-9 number buttons including the decimal key. By pressing the *Enter* key the current displayed number was saved and at the same time the display was cleared. The display itself provided participants to enter a 5-digit number with two digits after a separating decimal point as previously gathered numbers from infusion pump logs confirmed the validity of this number range [19].



Fig. 7. Experimental setup; serial interface in the *Medium* condition.

There were three versions of the interface used in the study. One where the number to be entered was displayed with a normal font-colour (black on white background) (group *Clear*), one where the number was displayed with a slightly lesser visible font-colour (light-grey on white background i.e. #BDBDBD) and the last one where the number to be entered was displayed in a hard to read faded font-colour (strong-faded grey on white background, i.e. #DFDFDF) in group *Obscured*.

In addition to the interface described above, which was accessed via logging into a website, the instructions as well as an informed consent document were printed out on paper for the participants to read and sign. Additionally, there was also a brief demographic questionnaire, which was implemented as a Google form. Again, this was not used in the analysis reported here.

The computer used was an Apple MacBook Pro 13" connected to an external keyboard and monitor with a 17" size display at 1920x1080 resolution. The browser used for the experiment was Mozilla Firefox 10 running under MacOSX 10.7.5 (Lion).

The interface ran in the browser, the browser window was maximised and controls were disabled from being shown so that only the interface area itself was visible. All other applications on the machine were closed and the auto-completion mode of the browser was deactivated. Furthermore, the keyboard was prepared so that participants could only use the numpad to enter numbers all other keys were disabled. Stickers were put in place on the '+' key to make it easier for the participants to associate the particular '*Clear*' key function.

Each number a participant had to enter, whether it was for the training to get familiar with the controls or the main study, was randomly generated in the game and displayed in the left upper corner of the interface (see Fig. 7). A smartphone was also used for this study to stop the total duration each participant used to enter as many numbers as possible.

3.5 Procedure

After a participant had read the instructions, he or she was asked to become familiar with the use of the keyboard controls and the display of the interface. For this purpose they were asked to run 10 training trials, which were separate from the main experiment. Participants were asked to sit straight in front of the monitor to avoid a possible effect on the perception of the numbers as the *Medium* and *Obscured* information could be perceived differently from a different angle. After a participant became familiar of how the serial number-entry interface worked he or she was then asked to proceed to the main study. Randomly selected participants (first come first serve basis) were asked to enter the given numbers either in the *Clear*, *Medium* or *Obscured* group within 5 minutes as quickly and as accurately as possible. The experimenter stopped the participant at the end of the 5-minute timeframe. During the experiment participants themselves could not see how much time had passed. Each time a participant pressed the *Enter*-key the number he or she just entered which was shown in the centre display was saved to the database and a new number was instantly and randomly generated at the upper left display. Additionally, at the same time the numbers in both displays were compared and if they matched the score were increased by 100 points. Any other entry than the correct number resulted in a decreasing of the score by 50 points. After finishing the number-entry task, the experimenter offered each participant a small break. Each participant was then asked to fill out a small online demographic questionnaire and the participants were debriefed as to the goals of the study.

3.6 Results

The hypothesis for this experiment was that there is a significant difference in the number of errors between the three conditions *Clear*, *Medium* and *Obscured*. Participants entered in total 2391 numbers for all three groups (see Fig. 8), which resulted in 44 total errors (1.8%) and 149 corrected errors. 849 numbers were entered in group *Clear* (28 errors), 796 in group *Medium* (12 errors) and 746 in group *Obscured* (4 errors), respectively. The mean of the errors made in the *Clear* group was 2.8 (sd = 2.20) whereas in the *Medium* group was 1.2 (sd = 1.39) and in the *Obscured* group 0.4 (sd = 0.52).

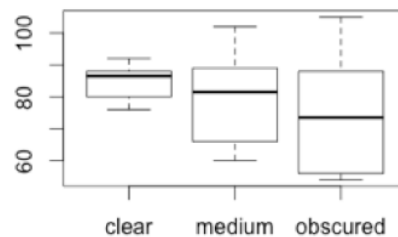


Fig. 8. Boxplot of amount of numbers entered by all participants in each condition.

A Kruskal-Wallis test was conducted, as error data does not follow a normal distribution (see Fig. 9). The results showed a highly significant difference between the errors made in the three groups ($H = 9.78$; $df = 2$; $p = 0.007$). However, there was no significant difference in the total numbers entered between the groups ($H = 2.08$; $df = 2$; $p = 0.35$) (see Fig. 8) and the number of corrected errors ($H = 0.23$; $df = 2$; $p = 0.89$). The mean of the numbers entered in the *Clear* group was 84.9 (st.dev = 5.45) whereas in the *Medium* group was 79.6 (st.dev = 13.95) and in the *Obscured* group 74.6 (st.dev = 17.61). Additionally, the mean of the corrected errors in the *Clear* group was 5.2 (st.dev = 5.18), whereas in the *Medium* group was 5.3 (st.dev = 6.80) and in the *Obscured* group 4.4 (st.dev = 3.59).

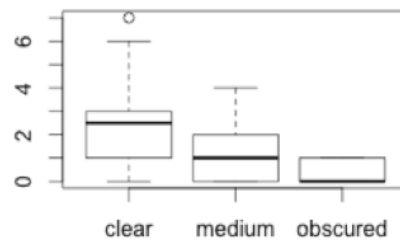


Fig. 9. Boxplot of the number of errors made by all participants in each condition.

Additionally, taking into account the number that each individual entered in relation to the errors, the rate of errors was also significantly different ($H = 8.89$; $df = 2$; $p = 0.01$). Moreover, the follow-up multiple comparison testing ($p = 0.05$) of the total errors made by each participant in each group revealed, that only the difference between group *Clear* and *Obscured* is significant (observed difference of 11.75 > critical difference of 9.43). A second follow-up test ($p = 0.05$) of the rate entered by each participant in each group revealed also that only the group *Clear* is significantly different from group *Obscured* (observed difference of 11.30 > critical difference of 9.43).

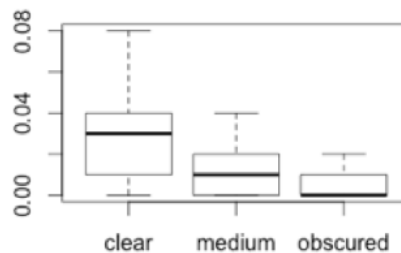


Fig. 10. Boxplot of the error-rate of each participant in each group.

Eleven participants (36%) made at least one error with only one person making no error in the *Clear* group, 4 persons making no error in the *Medium* group and 6 persons making no error in the *Obscured* group.

3.7 Discussion

Like the text-entry study the results showed that participants in the *Medium* and *Obscured* group achieved higher accuracy i.e. less errors than those in the *Clear* group. However, only the difference between *Clear* and *Obscured* is significant, which again suggests that the experimental manipulation similar to those suggested by the pilot has worked and there is a trend indicating that a more obscured representation of information can lead to less errors being made in data-transcription tasks.

Still, as observed in the pilot a presentation of content that is too obscure might have the opposite effect of making numbers simply too hard to read and therefore resulting in people abandoning the task.

Nevertheless, the discovered result is not due to possible speed-accuracy tradeoffs. Yes, participants entered fewer numbers in the *Obscured* group than in the other groups however not significantly different.

Interestingly, whilst the number of committed errors was fewer there is no difference in the number of the corrected errors. This may suggest that people use different strategies to correct errors or people are self-detecting their slip errors. While people who commit the errors think their action was correct, those who correct errors are aware of their mistakes suggesting that they are self-evaluating their actions on the fly. However, if people use a memory intensive strategy i.e. a strong encoding in memory then perhaps the corrected errors are made by comparison with memory rather than comparison with the original. It is interesting to note that there are fewer corrected errors, which also suggests different mechanisms behind corrected and uncorrected errors. Like other comparable studies [12, 18] the rate of number-entry errors is still low (1.8% in this case) but this does not mean it is less important to investigate why these errors happen. It is known that misprogrammed infusion pumps are the second most frequent cause of medication errors making it imperative to reduce as many errors as possible [16]. An error rate of 1.8% in medical device context can still result in grave consequences. And therefore, whilst the total number of errors is modest, there is still much to explore and to learn.

4 Overall discussion and Conclusions

We have presented two studies, which showed that less visible content could affect the information processing strategy adopted in a text-entry as well as in a number-entry task and, counter-intuitively, improved accuracy in the specific task.

This therefore suggests that less visible information can result less errors being committed in both text- and number-entry tasks. Moreover, the results show a strong support for the effect of *System 1* vs. *System 2* in the context of interaction. If tasks require a higher cost of access then evoking *System 1* vs. *System 2* might overcome the various problems presented by [8].

Further work employing eye-tracking measures is required to explore whether the information processing strategies adopted by those in the two groups result in differences in the number of fixations on, and the speed of reading of the target text. Specifically, it will be important to understand whether the enhanced performance of participants in the *Obscured* group is the result of slower more deliberate processes such as additional perceptual monitoring of the target text resulting in enhanced memory for the text or more careful execution of the motor movements required to enter the content. When focusing on design, our discovered results represent a potentially important contribution for the design of future medical devices.

However, there is the possibility that this effect will wear off over time. Future work will focus on a long-term study, where we can explore how long the discovered effect lasts in data-entry tasks.

Additionally, fatigue is another issue that needs to be examined in this context. There is evidence in the literature that, for example, nurses who are on extended work duty possess a significantly decreased level of vigilance [15]. The grade of frustration is another interesting aspect to explore. For example, if the visibility of the *Obscured* text is too low i.e. if the text is too hard to read then there is the potential that people will get too frustrated in entering numbers (or text) and simply lower their efforts towards a successful completion of the task or even abandon the task completely. In future studies, we are keen to consider how this effect can be applied further and explored to improve safety-critical systems such as infusion pumps.

Moreover, as healthcare personnel do not usually transcribe information from an electronic source but also from paper-based sources there is further room to explore the application of our findings in this context.

There is also the argument in the literature that text is processed differently in sense of string and individual characters [4, 9]. Additional work is required to explore the relations between these paradigms. For example, an alternative design for a text-entry study could focus on using content in welsh-language and let participants re-enter the text again in form of a game to investigate how people process text either in individual characters or as connected strings.

We believe our results revealed important findings based on a new approach to understanding the text-entry task. This may have significant implications for the design of tasks in safety-critical systems. We are therefore keen to continue to explore this seemingly fruitful research direction.

References

1. Abelson, R.P.: Statistics as principled argument. Lawrence Erlbaum Associates, Hillsdale NJ. (1995).
2. Alter, A.L., Oppenheimer, D.M.: Unity the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*. 13, 219–235 (2009).
3. Back, J., Brumby, D., Cox, A.: Locked-out Investigating the Effectiveness of System Lockouts to Reduce Errors in Routine Tasks. *Psychology*. 1, 3775–3780 (2010).
4. Bradshaw, J.L.: Three interrelated problems in reading: A review*. *Memory and Cognition*. 3, 123–134 (1975).
5. Diemand-Yauman, C., Oppenheimer, D.M., Vaughan, E.B.: Fortune favors the Bold (and the Italicized): Effects of disfluency on educational outcomes. *Cognition*. (2011).
6. Dijksterhuis, A.: System 3 thinking. Presented at the Cognitive Psychology Section Annual Conference Proceedings (2012).
7. Evans, J.S.B.T.: In two minds dual-process accounts of reasoning. *Trends in Cognitive Sciences*. 7, 454–459 (2003).
8. Gray, W.D., Sims, C.R., Schoelles, M.J., Fu, W.: The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review*. 113, 461–482 (2006).
9. James, K.H., James, T.W.: Letter processing in the visual system: Different activation patterns for single letters and strings. *Cognitive, Affective, and Behavioral Neuroscience*. 5, 452–466 (2005).
10. Kahneman, D.: Explorations of the Mind Intuition. UC Berkeley Graduate Council Lectures, Humanities. (2008).
11. Morgan, P.L., Patrick, J.: Designing Interfaces That Encourage a More Effortful Cognitive Strategy. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 54, 408–412 (2010).
12. Oladimeji, P., Thimbleby, H., Curzon, P., Iacovides, J., Cox, A.: Exploring Unlikely Errors using Video Games: An Example in Number Entry Research. In *Workshop on Safety-Critical Systems and Video Games: Contradictions and Commonalities*, held at *Fun and Games 2012*, Toulouse, France (2012).
13. Reber, A.S.: Implicit Learning and Tacit Knowledge. *Journal of Experimental Psychology: General*. 118, 219–235 (1989).
14. Salthouse, T.: Perceptual, cognitive, and motoric aspects of transcription typing. *Psychol. Bull.* 99, 303–319 (1986).
15. Scott, L., Rogers, A., Hwang, W.-T., Zhang, Y.: Effects of Critical Care Nurses' Work Hours on Vigilance and Patients' Safety. *American Journal of Critical Care*. 15, 30–37 (2006).
16. Smetzer, J., Cohen, M.: Preventing medication errors related to drug delivery devices. In M. R Cohen (Ed.) *Medication Errors*, 2nd Ed. (2006).
17. Stanovich, K.E., West, R.F.: Individual differences in reasoning: Implications for the rationality debate. *Behav. Brain Sci.* 23, 645–726 (2002).
18. Thimbleby, H., Cairns, P.: Reducing number entry errors solving a widespread, serious problem. *Journal of the Royal Society, Interface / the Royal Society*. 7, 1429–1439 (2010).
19. Wiseman, S., Cairns, P., Cox, A.: A Taxonomy of Number Entry Error. HCI2011: The 25th BCS Conference on Human-Computer Interaction. (2011).
20. Zhang, J., Shortliffe, E.H., Patel, V.L., Freed, M., Remington, R.: The Nature of Human Errors An Emerging Interdisciplinary Perspective. Proceedings of the 22th Annual Conference of the Cognitive Science Society (CogSci2000), (2000).