Advance Access publication on 2 November 2015
doi:10.1093/iwc/iwv034

The Effects of Perceptual Interference on Number-Entry Errors

FRANK SOBOCZENSKI*, MATTHEW HUDSON AND PAUL CAIRNS

Department of Computer Science, University of York, Deramore Lane, Heslington, York YO10 5GH, UK
**Corresponding author: fs596@york.ac.uk*

Safety critical number-entry tasks, such as programming syringe pumps, are common occurrences in healthcare settings. However, humans are prone to error and healthcare staff must often work in distracting and high stress environments. Previous work has developed a potential intervention that could reduce errors using a phenomenon called the Perceptual Interference Effect (PIE). In this paper, the numbers to be entered are presented in a hard-to-read form and experimental studies showed that this reduced the errors that people made. The aim of this paper is to investigate the robustness of this effect in the context of a distracting environment and therefore to begin to move towards investigating its efficacy in real-world settings. We report on an experiment, which uses auditory distractors alongside the PIE effect to investigate the effect on number-entry errors. The results are promising: the number of errors is significantly reduced by PIE and the rate of making errors is reduced though significance is marginal. Nonetheless, the results are suggestive that the PIE could work even in distracting circumstances though the isolation of this phenomenon in experimental studies is challenging. We explore the implications for future studies into this effect for eventual application in real-world contexts.

RESEARCH HIGHLIGHTS

- Introducing perceptual interference in number-entry tasks can improve information processing and reduce errors.
- People do make fewer errors where perceptual interference is present. This is reflected in the error rate although marginal.
- Perceptual interference is showing some robustness in distracting environments but further more focused studies are needed before it can be reliably deployed for number-entry tasks in safety critical domains.

Keywords: human-computer interaction (hci); empirical studies in hci; number-entry; perceptual interference; human error; disfluency

Received 31 January 2015; revised 14 August 2015; accepted 12 September 2015

1. INTRODUCTION

Safety critical number-entry tasks, such as programming syringe pumps, are common occurrences in healthcare settings. However, humans are prone to error and healthcare staff must often work in high stress and distracting environments. Soboczenski *et al*. (2013) found that disfluency in the presentation of a number can decrease errors in both the transcription of sentences and in number-entry tasks. Something as simple as making the characters harder to read by presenting them in light grey rather than standard black font produced a significant reduction in the number of errors without adversely affecting the speed of either task. Examples of the potential information processing benefits of disfluency are abundant (Alter *et al*., 2007; Alter and Oppenheimer, 2008, 2009; Corley *et al*., 2007; Diemand-Yauman *et al*., 2011; Hernandez and Preston, 2013; Soboczenski *et al*., 2013; Song and Schwarz, 2008) but the focus of this research has rarely been applied to safety critical applications, in which information processing can be a matter of life or death. This study further investigates the potential for disfluency to reduce number-entry error, with a specific focus on number-entry in syringe pump devices.

The phenomenon of disfluency leading to better information processing is known as the Perceptual Interference Effect (PIE) (Diemand-Yauman *et al*., 2011; Kühl *et al*., 2014). The PIE states that by introducing disfluency into perceived information (i.e. obscuring text or numbers in a different typeface or colour) leads to more effortful and deeper processing, better memory encoding as well as memory retrieval (Kahneman, 2008). The PIE can, therefore, lead to fewer errors being committed in transcribing previously perceived information. A transcription task can be in the form of text transcription or number transcription in this context.

Previous research applying the PIE to transcription was done in the context of formal experimental studies but our long-term goal is to reduce transcription errors in real-world, safety-critical tasks, specifically number-entry for medical devices. Therefore, much still needs to be done to understand the robustness of the PIE as a practical intervention. This paper aims to extend the previous work by investigating whether the PIE can still reduce errors in the presence of auditory distractors though this has still been done as a formal experiment. The results of this study show that, similar to Soboczenski *et al*. (2013), disfluency does not have a significant effect on number-entry speed, but does reduce the number of errors made. However, the change in error rate is only marginally significant. Thus, the results are, at best, suggestive of the possible robustness of the PIE. At the same time, we have to acknowledge the difficulties of isolating number-entry errors in experiments so that they may be carefully studied. This leads to further consideration of how to develop insightful experiments into the PIE for safety-critical domains.

2. LITERATURE REVIEW

2.1. Errors

Errors are part of human nature and constantly appear in daily tasks. From dialling the wrong phone number to typos, people make errors in all kinds of situations, even in tasks with the lowest level of complexity (Kantowitz and Sorkin, 1983). The online repository *Errordiary* is a collection of such everyday errors. *Errodiary*, which can be contributed to by using the hashtag *#errordiary* on Twitter, was created to increase the discussion about human error and its causes (Wiseman *et al*., 2011).

Although many of the errors are little more than an inconvenience, in safety-critical domains such as aviation, nuclear systems or healthcare, a small error can have severe consequences. In these domains, the problem is not trivial, in the aviation domain alone an analysis of the flight recorders revealed that 70–80% of accidents are based on human error or based on a chain of failures in relation to the human factor (FAA, 2010; Martins *et al*., 2013). The literature provides us with numerous cases of human error, some resulting in tragedies (Casey, 1998, 2006; Kostopoulou and Delaney, 2007).

Research around number-entry errors is a relatively young but fast expanding area, and it is only recently that studies have shown that people make number-entry errors frequently and that many systems do little to detect or mitigate errors (Thimbleby and Cairns, 2010). 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. In healthcare, the process of studying errors is only now emerging (Wiseman *et al*., 2011).

2.2. Healthcare errors

Healthcare is an increasingly complex, data-intensive and technology-intensive domain. Healthcare workers are constantly interacting with a variety of systems under physical and cognitive stress. It is a highly demanding area focused on efficiency, accuracy and effectiveness. While safety-critical domains such as aviation have greatly benefited from human error research, in healthcare many errors are not reported. However, Kohn *et al*. (2000) stated that in the USA more people die from medical errors than from road accidents, breast cancer or AIDS. While errors in healthcare are prevalent, the majority of errors are not caused by people acting carelessly but by erroneous systems and procedures which lead people to commit errors (Kohn *et al*., 2000).

The healthcare domain is increasingly reliant on a vast amount of technology. This includes ordinary administrative equipment and complex medical devices in the operating room such as respiratory ventilators, defibrillators or infusion pumps. The act of using a syringe pump contains a range of steps, which could ultimately lead to a critical error caused by inaccurate information processing. Transcription errors can occur due to prescriptions or dosages being noted down using error-prone abbreviations (Gaunt and Cohon, 2007), slips such as decimal point or keystroke errors during dosage calculation can cause fatalities (Thimbleby, 2007). Errors can also be caused by the lack of training or upkeep of skills on a particular device, cognitive slips, capture errors, mode errors or by inherently unintuitive, misleading or dangerous design (Obradovich and Woods, 1996; Thimbleby and Cairns, 2010; Vicente *et al*., 2003; Wallace and Kuhn, 2001; Wiklund, 1995; Wiseman *et al*., 2011; Zhang *et al*., 2003). The United States Food and Drug Administration (FDA) and other official bodies, such as the Institute for Safe Medication Practices (ISMP) are recognizing the dangers inherent with medical devices and published guidelines to address as many issues as possible (FDA, 2000).

While double-checking drug dispensing information, calculations and device input can detect errors, in some situations, double-checking can be rendered ineffective. If a double-check is not done by a second individual, then the usefulness is reduced, as it is more difficult to find one's own errors due to confirmation bias (Grissinger, 2003). Even if a double-check is performed by a second individual, it can still be rendered ineffective if it is not conducted independently (Grissinger, 2003), or if the double-checker is distracted (ISMP, 2007). It must also be stated that while training is important, it cannot help to prevent slip errors as described in Byrne and Davis (2006) and Back *et al*. (2010).

2.3. Disfluency: the PIE

Disfluency theory is the idea that applying perceptual difficulties to data will trigger deeper cognitive processing (Alter *et al*., 2007), and therefore can lead to higher learning outcomes (Diemand-Yauman *et al*., 2011). Disfluency has been shown to affect the way humans perceive information. Hernandez and Preston (2013) found that disfluency reduces confirmation bias, and Alter and Oppenheimer (2008) state that disfluency encourages abstract interpretation.

These somewhat counter-intuitive findings can be conceptualized in terms of the System 1 and System2 theory of cognition (Kahneman, 2008; Stanovich and West, 2002). Kahneman (2011) bases a theory of thought, intuition and skill acquisition on the concept of two systems within the mind. These metaphorical agents within the mind are called System 1 and System 2. System 1 deals with automatic operations with a low cognitive cost, for example when we look at a spoon, we known instantly that this is a spoon. This instant recognition is System 1 at work. System 2 refers to our controlled mental operations, when we stop to consider and analyse the world around us. Kahneman (2011) refers to these systems as thinking fast (System 1) and thinking slow (System 2). In terms of training, we can think of learning as moving a skill from something, which must be done using System 2, to something which can be done using System 1. Kahneman (2011) uses the example of learning to read, when we begin to learn we must work hard to recognize individual letters and assemble them into syllables and words, requiring System 2. After practice, this process moves into System 1, becomes automatic, and so letters and syllables are instantly recognizable. This use of System 1 frees up capacity in the mind so that System 2 can then concentrate on the next level of skill required for reading.

Alter *et al*. (2007) suggest that disfluency serves as an alarm to the mind which activates more analytical reasoning (System 2), and Alter and Oppenheimer (2009) argue that disfluency encourages more effortful information processing rather than intuitive automatic reading. Alter *et al*. (2007) state that if ‘information is processed easily or fluently, intuitive (System 1) processes will guide judgment. If information is processed with difficulty or disfluently, however, this experience will serve as a cue that the task is difficult or that intuitive response is likely to be wrong, thereby activating more elaborate (System 2) processing’. The concept that ‘metacognitive experiences of difficulty’ change the way humans think is not just a psychological theory, but has been shown to activate parts of the brain which are responsible for deliberative and effortful thought (Alter *et al*., 2007; Boksman *et al*., 2005; Botvinick *et al*., 2001).

In terms of information processing, disfluency has been shown to aid in the learning and transcription of information. Diemand-Yauman *et al*. (2011) argue that information represented in a less readable format, for example a harder-to-read font, can be more accurately memorized. This is known as PIE. In their studies, they made fonts harder to read by using different typefaces with different weights, *italicization* and also using 60% grey scale.

Soboczenski *et al*. (2013) applied the PIE to the task of number transcription on the basis that transcription requires the processes of accurately (albeit briefly) recognizing and mentally encoding a string of numerals to-be-transcribed before making the motor movements to convey those numerals into a device (Salthouse and Saults, 1987). Disfluency could aid in the accurate encoding of the numerals by overriding the automatic but possibly inaccurate System 1 and engaging System 2 in more effortful reading and therefore more accurate encoding. In their studies, information was made harder to read only by presenting it in a light grey rather than black font, and it was found that this disfluency did indeed lead to fewer errors in both sentence transcription and number-entry tasks.

Despite these unexpected results in support of the PIE, the effects of disfluency have yet to be exhaustively explored and Eitel *et al*. (2014) found that the benefits of disfluency are not entirely consistent, perhaps depending on individual differences and context. Kühl *et al*. (2014) call for caution in making generalizations regarding disfluency and advocate the need for more research into the topic.

2.4. Auditory distractors

In order to examine the potential of the PIE as a way to reduce number-entry errors in safety-critical domains, we wished to consider the robustness of the PIE in a distracting environment, such as might be found on hospital wards. Though a simple task, number-entry still requires focus and attention and distractors could move attention away from this task leading to further errors. Though the PIE has been seen to be effective in what might be considered ideal conditions, it may be that there is no benefit from it in environments when there are competing demands on attention.

There has been substantial work in psychology to understand the nature of attention in particular how attention is captured by distractors. This has centred around the debate of whether distractors are filtered out shortly after perception (early selection) or they are processed fully but prevented from influencing actions or memory (late selection). The perceptual load theory of attention (Lavie, 2005) says that whether early or late selection happens depends on the perceptual load that a person is experiencing while doing their primary task. Perceptual load has mostly been studied in visual search tasks. In this context, where there is low perceptual load, that is, few items to search through or items that are easily distinguished from the target item, there is a lot of spare capacity for processing distractors and so late selection occurs. In high perceptual load, that is many items or items that are easily confused with the target, there is much less capacity for processing distractors and so early selection occurs. The upshot is that in situations with high perceptual load, distractors are less distracting.

In the context of studying the PIE, it could be the case that presenting the numbers to-be-transcribed in a hard-to-read format could itself be a source of perceptual load, thus leading to less susceptibility to distraction. However, this is not the case. Lavie and De Fockert (2003) made their stimuli harder to read by reducing contrast between the text and the black background and also by reducing the size of the text. They examined the effect of this on the processing of visual distractors during a visual search task. It was found that there was no effect of the text format on susceptibility to distractors. Their study has similarities to ours because reduced contrast is used to provoke the PIE but we are using a more interactive task, number-entry, rather than visual search for a target and also we use auditory distractors rather than visual distractors. Nonetheless, the results from Lavie and Fockert suggest that the stimuli we use are not simply influencing perceptual load and therefore influencing awareness of distractors through that route. In many workplace situations, distractions do not come from visual distractions but rather the auditory landscape around people. Few workplaces are totally silent and hospital wards are no different. We therefore decided to use auditory distractors as having some comparison, admittedly idealized, to the situation of number-entry on a medical device in a busy ward. Hearing seems to play a role in generally monitoring the environment and thus can act as an 'early warning' system with unexpected sounds capturing attention even if they are irrelevant to the task in hand (Dalton and Lavie, 2004). Murphy *et al.* (2013) also showed that unlike the visual search task where perceptual load can result in the filtering out of distractors, in an auditory search task, distractors are still perceived regardless of the auditory perceptual load. This has also been seen cross-modally with the perceptual load of a visual search task having no effect on the perception of auditory distractors (Tellinghuisen and Nowak, 2003). In particular, in the visual search task of looking for either an X or N in a group of letters, an auditory distractor of 'X' or 'N' either increased or decreased reaction times depending on whether the distractor was the same or different from the target letter.

At the same time, there is some evidence that increased engagement with a task can reduce the effect of auditory distractors. Of particular relevance to the work in this paper is that of Halin *et al.* (2014a; 2014b). They use the PIE to examine how background speech could influence people's abilities to engage in proofreading tasks (Halin *et al.*, 2014a) and recall tasks (Halin *et al.*, 2014b). In the easy-to-read conditions, participants perform less well with the auditory distraction but in the hard-to-read condition, people perform equally well. This may because speech engages complex processes and so the PIE effect interferes with some of those processes to reduce the effectiveness of speech as a distraction. Also, though this work shows promise for the robustness of the PIE, it should be noted that they contrasted the situation of auditory distraction with silence as opposed to relevant versus irrelevant distractors. However, similar effects are seen in more 'pure' tasks (Sörqvist, 2014) inattentional deafness (Macdonald and Lavie, 2011): people remain unaware of changes to the auditory environment depending on the visual perceptual load they are experiencing. The stimuli used in these experiments are much simpler and comparable to the visual search tasks discussed above. Moreover, they compare irrelevant to relevant distractors. The role of distractors though, as a changing soundscape, is very different from the effect of general background speech as a distraction.

Thus, auditory distractions have an analogue to the noisy, distracting environments that motivate our work in this area. However, the influence of auditory distractions in relation to the PIE as applied in the number-entry task is unknown. The PIE may enable better focusing on the number-entry task and hence reduce errors or it may be that, for this task, it is unable to block the effect of distractors.

2.5. Summary

In conclusion, research in number-entry is a relatively young field. The research community has only recently focused on reducing errors in number-entry applications. Healthcare is one of the safety critical domains, which experiences a constant rise and change of technology, which relies heavily on accurate interaction, i.e. text or number transcriptions. Traditional approaches such as double entry checking have been proved to not always be reliable safety measures in number-entry. Yet disfluency theory might present a simple albeit counterintuitive approach to reduce such errors. The PIE should introduce perceptual difficulties to data, which triggers deeper cognitive processing (Alter *et al.*, 2007), therefore lead to better encoding and retrieval of the data. This effect has been shown in a laboratory setting to result in higher accuracy and fewer number-entry errors (Soboczenski *et al.*, 2013). However, little is known about how the effect behaves in real-world applications in particular in the presence of auditory distractions. The existing psychology literature suggests the potential of the PIE to overcome auditory distractions but the specific effect in the context of number-entry is unknown.

3. EXPERIMENT

The aim of this study was to reproduce the PIE in a number-entry task in the presence of auditory distractors. The distraction in this study consists of an audio track, which was played to participants as they took part in a number-entry task. This study introduces two distinct audio tracks (numerals and letters) as distractions in two visibility conditions (*Clear* and *Obscured*) to investigate the robustness of the PIE in terms of memory encoding and retrieval. An audio track of numerals is intended to present a greater distraction in the number-entry task than letters as the entry of numbers into a device requires the participants to encode and then enter numerals. Hearing letters ought not to interfere with this, following Tellinghuisen and Nowak (2003). The expectation is therefore that numeral distractors will provoke more errors than letter distractors. However, if the PIE is robust, it should still be seen in both the numeral and letter conditions, which means that errors should be reduced under both conditions.

H1: The PIE can be reproduced in the number-entry task. Hence there would be a significant difference in the number of errors between the *Obscured* and *Clear* conditions.

H2: Numeral distractors lead to more errors than the letter distractors.

3.1. Design

A between-subjects design was used for this experiment. The study, had the following four conditions (as a result of two independent variables): the grade of visibility in which the information to-be-transcribed was presented in either black (Red–Green–Blue colour-code: 0,0,0) or grey (RGB-code: 223,223,223), and the type of audio track participants had to listen to while transcribing numbers (Letters or Numerals).

Clear-Letters (*CL*): in this group, participants saw the numbers to-be-transcribed in a clear black font while listening to an audio track speaking single letters from the alphabet.

Clear-Numbers (*CN*): in this condition, people saw the to-be-transcribed information in a clear black font while listening to an audio track speaking single numerals.

Obscured-Letters (*OL*): participants in this condition saw the to-be-transcribed numbers in a grey colour font while listening to the single letters audio track.

And in the *Obscured-Numbers* (*ON*) condition, participants listened to the numeral audio track while transcribing information presented in a grey coloured font.

The dependent variables were the total numbers entered, the total number of errors made, including the rate of errors and also the number of corrected errors in each condition.

3.2. Participants

A total of 40 participants (17 women) were randomly recruited by personal invitation in the Morrell library of the University of York. There were 10 participants in each group. The sample size was considered adequate as a previous study in Soboczenski *et al*. (2013) showed a highly significant effect in a similar experimental set-up (30 participants, 3 groups with 10 people in each group, respectively). All of the participants were students at the University of York in different departments, except one sales administrator. Participants were between 19 and 29 years old (mean 24.25; standard deviation 5.12). Five participants had previous experience in the healthcare sector. Among those were two nursing students. Four participants stated that they never used a tablet before, one in the *Clear-Letters* condition, one in the *Obscured-Numerals* condition and two in the *Obscured-Letters* condition. None of the participants stated that they never used a touch-based device. All of the people stated that they had normal vision and normal hearing. One participant in each condition stated that they did not enjoy the task and only one in the *Obscured-Letters* condition stated that the task was difficult.

3.3. Number-entry tasks and interface

The tasks for this study were framed as a number-entry game on the iPad to enter as many numbers as quickly and as accurately as possible in a calculator-based interface. The number-entry task was designed as a game with a score where participants could evaluate themselves on how well they perform while entering numbers. The score would increase by 100 for each correct number a person entered and decrease by 100 for a wrong number or for accidentally pressing the enter key. The score was introduced to motivate participants to enter as many numbers as possible in the short timeframe and to increase the data pool for analysis. Participants had to enter randomly generated numbers in a range of 0–999.99 (5% without decimal places). The font-type used to represent the numbers was System Bold in font-size 49. The number-entry interface did not provide any cue as to a potential error (e.g. double taps on decimals) other than the visible transcribed number in the target display. However, the log file provided detailed insight into timestamps and all button presses. The application always used a white background. Participants had a time limit of 5 min.

3.4. Materials

Participants were given paper-based instructions and an informed consent form to read and to sign before proceeding to the training and the main task. A Google demographic questionnaire was given to participants after the study. The research instrument used for this study was an iPad2 running iOS 6.3.5. Additionally, this experiment required participants to listen to one of two audio tracks: one with random numerals and another one with random letters (see Fig. 1).

The audio track consisted of a string of numerals or letters. The audio track lasted for as long as the experiment ran (5 min) and with each letter or numeral being read out with a one-second gap. Both tracks were set up as an MS

Figure 1. Experimental set-up: iPad running the number-entry app in the clear condition with the number to-be-transcribed in the upper left corner and the target area in the centre above the number-pad.

Word document and then read by the inbuilt voice of the used Macbook Pro. For this purpose, Apple's Samantha voice package was downloaded and installed prior to the study as this voice showed the closest resemblance to a human voice. The numerals read to participants were random single numerals from 0 to 9. The letters were random single letters from the whole alphabet.

Unlike the experiments tailored to the study of distraction in either visual or auditory search, it was not possible to tightly synchronize the timing of the auditory distractors to the number-entry task. This was because of the difference between a visual search task and a number-entry task. In a visual search task, the participant simply has to respond to having found the target. In a number-entry task, the participant needs to make a series of key presses and that may depend on the participants' typing speed as well as any of the intended manipulations. The headset used was a Somic Stereo ST-908 and a smartphone was used to manage the time.

3.5. Procedure

The experiment was conducted in a controlled study room in the library. All participants were run in individual sessions, which were managed by personal invitation. Each participant was first welcomed and then given a copy of the informed consent form and the instructions to read and to sign. This was then followed by participants entering six training numbers while listening to the specific audio track (either Letters or Numbers) (Fig. 1).

Participants were then asked to enter the given numbers in one of the four conditions within 5 minutes as quickly as possible. The experimenter stopped the participant after the 5 minutes were up. After finishing the main number-entry task, participants were asked to complete the demographic questions followed by the debriefing of the study.

3.6. Results

The data analysis in this study is of non-parametric data (error-data). As this analysis is looking for two-way effects, a (2×2) *Analysis of Variance* (*ANOVA*) is the suitable test, as there is no well-established test for non-parametric data. However, main effects where they are seen in the ANOVA are confirmed with a non-parametric Mann–Whitney test.

A total of 3680 numbers were entered during the study, which resulted in 66 total errors (1.8%). 957 numbers were entered in condition *Clear-Letters* (19 errors), 884 in condition *Clear-Numbers* (23 errors), 914 in condition *Obscured-Letters* (14 errors) and 925 in condition *Obscured-Numbers* (10 errors). Table 1 illustrates the relevant descriptives in relation to the total numbers entered in each condition.

An alpha level of 0.05 was used for all statistical tests. The data were analysed using a two-way (2×2) ANOVA for a between-participants design, with the grade of the introduced disfluency, i.e. visibility and the type of audio distraction as unrelated sample variables. In regard to the numbers entered in total, there was no main effect of visibility, $F(1,36) = 0.0$; $P = 0.99$; $\eta_p^2 < 0.001$, with those in the *Obscured* conditions transcribing slightly more numbers than in the Clear conditions but not significantly. There was also no main effect for audio $F(1,36) = 0.309$; $P = 0.58$; $\eta_p^2 < 0.001$, with a few more numbers entered in the letters conditions but again not significantly more than in the numbers conditions. There was also no interaction effect (Visibility × Audio) $F(1,36) = 0.568$; $P = 0.46$; $\eta_p^2 = 0.016$ (see Fig. 2).

In regard to the total errors made, there was a marginal main effect of visibility, $F(1,36) = 4.084$; $P = 0.051$; $\eta_p^2 = 0.10$, with those in the *Obscured* conditions making marginally fewer errors than participants in the *Clear* conditions. This main effect was confirmed by the Mann–Whitney test ($P = 0.04$; $W = 272.5$). There was however no main effect for audio $F(1,36) = 0.0$; $P = 1.0$; $\eta_p^2 < 0.001$, with participants in both audio conditions making about the same amount of errors. There was no interaction effect (Visibility × Audio) $F(1,36) = 0.807$; $P = 0.38$; $\eta_p^2 = 0.022$ (see Fig. 3).

Table 2 illustrates the descriptives regarding the total errors made by condition.

Table 1. Descriptives: mean (SD) for the total numbers entered.

		Visiblity	
		Clear	Obscured
Audio	Letters	95.7 (21.81)	91.4 (16.49)
	Numbers	88.4 (17.90)	92.5 (13.20)

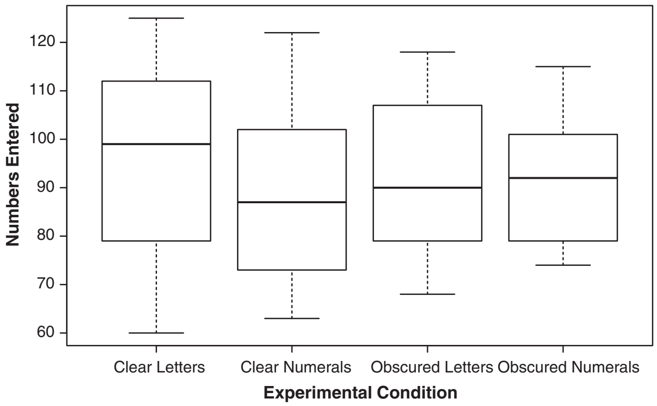

Figure 2. Boxplot of total numbers entered by each participant in each condition.

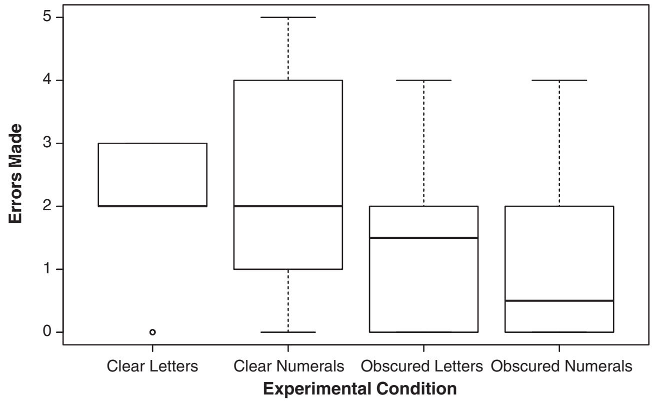

Figure 3. Boxplot of total errors made by each participant in each condition.

Table 2. Descriptives: mean (SD) for the total errors made.

		Visiblity	
		Clear	Obscured
Audio	Letters	1.9 (1.1)	1.4 (1.43)
	Numbers	2.3 (1.7)	1.0 (1.33)

Thirteen participants made no errors during the entire study. Two participants made no errors in the *Clear-Letters* condition, two participants made no errors in the *Clear-Numbers* condition, four participants made no errors in both the *Obscured-Letters* condition whereas five participants made no errors in the *Obscured-Numbers* condition (68% of the participants made at least one error). The highest error-count was recorded in the *Clear-Numbers* condition.

In regard to the error-rate, there was a marginal main effect of visibility, $F(1,36)=3.70$; $P=0.062$; $\eta_p^2=0.093$. There was no main effect for audio $F(1,36)=0.005$; $P=0.945$; $\eta_p^2<0.001$ and no interaction effect $F(1,36)=1.18$; $P=0.285$; $\eta_p^2=0.032$ (see Fig. 4). A Mann–Whitney test confirmed the marginal effect of visibility on error rate ($W=264.5$, $P=0.078$).

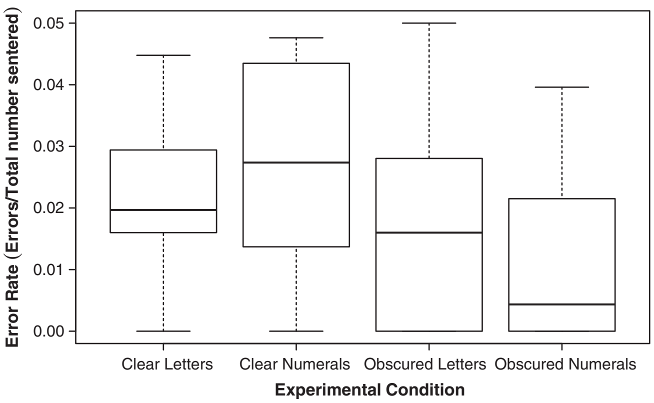

Figure 4. Boxplot of the error rate by each participant in each condition.

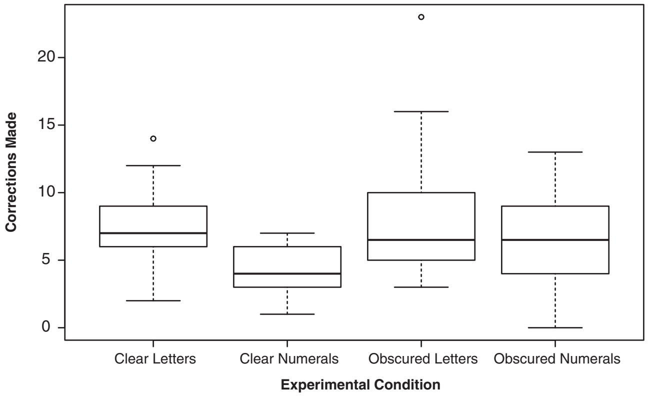

Figure 5. Boxplot of the total corrections made by each participant in each condition.

Participants made in total 268 corrections during the study of which 75 corrections were made in condition *Clear-Letters*, 42 in condition *Clear-Numbers*, 90 in condition *Obscured-Letters* and 61 in condition *Obscured-Numbers*. The highest amount of corrections made by a participant was 23 (resulting in 0 errors) in the *Obscured-Letters* condition.

In regard to the total corrections made, there was no main effect of visibility, $F(1,36) = 1.565$; $P = 0.22$; $\eta_p^2 = 0.042$, with those in the *Obscured* conditions making more corrections than participants in the *Clear* conditions but not significantly more. There was, however, a main effect for audio $F(1,36) = 5.202$; $P = 0.029$; $\eta_p^2 = 0.13$, with participants making significantly more corrections in the letters conditions than the numbers conditions, which the Mann–Whitney test confirmed ($P = 0.04$; $W = 275$). There was also no interaction effect (Visibility × Audio) $F(1,36) = 0.022$; $P = 0.88$; $\eta_p^2 < 0.001$ (see Fig. 5). An identical analysis of the rate of corrections against the total amount of numbers entered gives exactly the same picture (Table 3).

Table 4 presents a summary of the errors participants made classified according to Wiseman *et al*. (2011) and structured in the four conditions.

4. DISCUSSION AND CONCLUSIONS

As in previous work (Soboczenski *et al*., 2013), this study aimed to replicate the PIE in a number-transcription task but to investigate its robustness in the presence of audio distractors. The total number of errors made was reduced by the PIE and the error rate was also marginally reduced as well. Moreover, this is not a speed-accuracy trade-off with participants slowing down in order to be more accurate. This reflects the results of (Soboczenski *et al*., 2013) though they are not so unequivocal. Of course, the main difference is that the current task was done in the presence of an audio distraction. Though the intention was that the letters would be less distracting than the numerals, this does not seem to have come through in the measures used. There are no main effects of audio distractors on either the total errors made or the error rate. A possible explanation could be that someway the numeric distractors also contributed to the PIE by increasing focus and concentration on the task. This would mean that the numeric distractors were not sufficiently distracting enough. There are indications in the literature that people process distractors depending on the type of processing demands. Specifically, Lavie and De Fockert (2003) found that higher perceptual load stimuli in a ‘relevant’ task reduced the processing of ‘irrelevant’ distractors. In general, this could be transferred to our study. It seemed that numeric distractors are somewhat task relevant to numerical entry whereas letters are not which is reflected in our hypothesis so the implication would be that task is not sufficiently demanding so that all distractors are being processed equally. However, Lavie and De Fockert (2003) were limited to the visual domain. In the mixed modality domain, Tellinghuisen and Nowak (2003) found no reduction in the effect of distractors depending on visual load but even so the effect of distractors was dependent on the relevance of the distractors to the task in hand. This is therefore an interesting feature contrasting with the existing literature that encourages us to further exploration but cannot be examined in any more detail from the data here as it was not the primary focus of this study. If it is therefore the case that letters and numerals are equally distracting for the number-entry task, then the spoken numerals are not more relevant than the spoken letters. This may be because numbers can be multiply represented by people as strings of digits, spoken words or spatially (Dehaene, 1997) and the spoken representation of numerals is not used in the number-entry task. Alternatively, it might be that all auditory distractors are equally intrusive for this task but this might perhaps have led to a higher error rate overall than previous studies, which was not the case. In fact, given the generally lower error rate, it may be that there was some increased engagement as a consequence of the distracting audio and the PIE could only marginally improve upon this.

The only significant difference due to distractors was in corrected errors with participants making significantly more corrections (both totally and as a rate of corrections) when listening to the letters distractors. If anything, this suggests that letters were more distracting than numerals but in such a way that participants were able to catch the errors for themselves. We tenuously suggest that it might be a result of interference between the motor memory to type a letter with the intention to type a particular digit (Salthouse and Saults, 1987) or perhaps an indication that letters and numerals are processed differently (Dehaene, 1997). But really this is essentially a surprising result that would need substantially more investigation before we could provide a more insightful explanation.

Table 3. Descriptives: mean (SD) for the total corrections made.

		Visiblity	
		Clear	Obscured
Audio	Letters	7.5 (3.7)	9.0 (6.2)
	Numbers	4.2 (2.0)	6.1 (4.2)

Table 4. Summary of the errors made in all conditions.

	Occurrence in condition				Example
Type	*Clear-Letters*	*Clear-Numerals*	*Obscured-Letters*	*Obscured-Numerals*	(to-be-entered → entered)
Out-by-factor-of-10	4	4	2	3	658.71 → 65.87
Transposition error	2	3	6	3	289.32 → 298.32
Wrong digit	10	10	5	4	43.02 → 43.03
Wrong number	1	1	1	0	186.47 → 0.81
Extra digit	1	0	0	0	548.81 → 5483.81
Missing digit	0	5	0	0	81.04 → 8.04
Early termination	1	0	0	0	79.91 → 79.9

Table 4 shows a summary of the different types of errors that participants made structured according to previous work in the field (Wiseman *et al*., 2011). Transposition errors seem to happen across the conditions indicating possible motor skill dependence (Salthouse and Saults, 1987), whereas missing digit errors only occurred in one of the clear conditions indicating what could be seen as possible encoding errors. Similarly, wrong digit errors also seemed to occur more in both clear conditions. However, due to the few number of occurrences of errors, these differences can only be seen as indicative but not conclusive. These could, however, be used as working hypotheses to consider in future studies.

Overall, then, it seems there is a potential effect of the PIE in improving number-entry accuracy, even in the presence of auditory distractors. The effect size is modest and only marginally significant but bearing in mind the healthcare context where number-entry is a widespread and safety-critical task, even small effects of this sort can be important. However, we are some way from definitively establishing the robustness of the PIE. In particular, that the numeral distractors were not more distracting than the letter distractors is surprising and it may even be the case that any auditory distraction for this task can be effectively screened out. It may indeed be the case that when it comes to the number-entry task, it brings its own focusing effect that is therefore naturally resistant to distraction. It would be worth comparing this with other interactive tasks to see if it is possible to perceive the effectiveness of distractors in this context.

We have presented a study that successfully replicated the PIE and further investigated its robustness once levels of distractions are present. The results in this work suggest that the PIE has the potential to provide real-world benefits in safety critical domains, for example by obscuring drug prescriptions to promote more careful number-entry into syringe pumps. However, a real-world application of disfluency relies on the results of further investigation into the long-term use of the technique, and the potential interplay between disfluency, real-world distractions and existing protocols. Further investigation is needed to establish if the effects of disfluency change over time, i.e. do people become immune to its effects? Additionally, further studies need to focus on identifying clear and effective distractors for the task and demonstrate the PIE in this context as well as including a multi-tasking environment rather than straight distractions to move closer to a real-world ward environment. In terms of real-world distractions and existing safety protocols, studies, which provide analogous distractions, would be useful to establish the interplay between these concepts.

At this moment, there is no real-world implementation in safety-critical systems of the findings presented here or the PIE in general. Additionally, current PIE literature only focuses on reproducing the effect (Alter and Oppenheimer, 2009; Corley *et al*., 2007; Diemand-Yauman *et al*., 2011). We on the contrary showed over several studies that the PIE can be reproduced/applied in transcription tasks (Soboczenski *et al*., 2013). Moreover, we went one step further in presenting the first study with the aim to move the PIE into real-world applications in safety-critical environments. Yet there are still many questions about how the PIE reacts in such an environment, a concern that is reflected in the current literature (Eitel *et al*., 2014; Kühl *et al*., 2014). Future studies need to address this gap to gain a more complete understanding of the PIE and its impact on error reduction in transcription tasks. Further to that, we would like to raise awareness of the potential benefits for interface designers and device manufacturers by applying the PIE in the healthcare domain. The PIE represents a small and cheap manipulation, i.e. a small change in the presentation of the to-be-transcribed content with a potentially huge impact on safety in terms of accuracy.

REFERENCES

Alter, A.L. and Oppenheimer, D.M. (2008) Effects of fluency on psychological distance and mental construal (or why New York is a large city, but New York is a civilized jungle). Psychol. Sci., 19, 161–167.

Alter, A.L. and Oppenheimer, D.M. (2009) Unity the tribes of fluency to form a metacognitive nation. Personality Soc. Psychol. Rev., 13, 219–235.

Alter, A.L., Oppenheimer, D.M., Epley, N. and Eyre, R.N. (2007) Overcoming intuition: metacognitive difficulty activates analytic reasoning. J. Exp. Psychol.: Gen., 136, 569.

Back, J., Brumby, D. and Cox, A. (2010) Locked-out investigating the effectiveness of system lockouts to reduce errors in routine tasks. Psychology, 1, 3775–3780.

Boksman, K., Théberge, J., Williamson, P., Drost, D.J., Malla, A., Densmore, M., Takhar, J., Pavlosky, W., Menon, R.S. and Neufeld, R.W.J. (2005) A 4.0-T fMRI study of brain connectivity during word fluency in first-episode schizophrenia. Schizophrenia Res., 75, 247–263.

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S. and Cohen, J.D. (2001) Conflict monitoring and cognitive control. Psychol. Rev., 108, 624.

Byrne, M.D. and Davis, E. (2006) Task structure and postcompletion error in the execution of a routine procedure. Hum. Factors, 48, 627–638.

Casey, S.M. (1998) Set Phasers on Stun: And Other True Tales of Design, Technology, and Human Error. Aegean Publishing Company, Santa Barbara, CA.

Casey, S.M. (2006) The Atomic Chef: And Other True Tales of Design, Technology, and Human Error. Aegean Publishing Company, Santa Barbara, CA.

Corley, M., MacGregor, L.J. and Donaldson, D.I. (2007) It's the way that you, er, say it: hesitations in speech affect language comprehension. Cognition, 105, 658–668.

Dalton, P. and Lavie, N. (2004) Auditory attentional capture: effects of singleton distractor sounds. J. Exp. Psychol.: Hum. Percept. Perform., 30, 180.

Dehaene, S. (1997) The Number Sense: How the Mind Creates Mathematics. Oxford University Press, Oxford.

Diemand-Yauman, C., Oppenheimer, D.M. and Vaughan, E.B. (2011) Fortune favors the bold (and the Italicized): effects of disfluency on educational outcomes. Cognition, 118, 111–115.

Eitel, A., Kühl, T., Scheiter, K. and Gerjets, P. (2014) Disfluency meets cognitive load in multimedia learning: does harder-to-read mean better-to-understand? Appl. Cogn. Psychol.

FDA (2010) Federal Aviation Administration. DOT/FAA/AM-10/13, Office of Aerospace Medicine, Causes of General Aviation Accidents and Incidents: Analysis Using NASA Aviation. Safety Reporting System Data.

FAA (2000) Food and Drug Administration. Medical Device Use-Safety: Incorporating Human Factors Engineering into Risk Management.

Gaunt, M.J. and Cohon, M.R. (2007) Medication Errors 2. American Pharmacists Association, Washington, DC.

Grissinger, M. (2003) The Virtues of Independent Double-Checks: They Really Are Worth Your Time! Institute for Safe Medication Practices (March).

Halin, N., Marsh, J.E., Haga, A., Holmgren, M. and Sörqvist, P. (2014a) Effects of speech on proofreading: can task-engagement manipulations shield against distraction? J. Exp. Psychol.: Appl., 20, 69.

Halin, N., Marsh, J.E., Hellman, A., Hellström, I. and Sörqvist, P. (2014b) A shield against distraction. J. Appl. Res. Mem. Cognition, 3, 31–36.

Hernandez, I. and Preston, J.L. (2013) Disfluency disrupts the confirmation bias. J. Exp. Soc. Psychol., 49, 178–182.

Institute for Safe Medication Practices Canada ISMP. (2007) Fluorouracil incident root cause analysis. Final draft, April 5, 2007; final report, April 30, 2007; online May 22, 2007. www.ismp-canada.org/download/reports/FluorouracilIncidentMay2007.pdf (accessed May 5, 2011).

Kahneman, D. (2008) Explorations of the Mind. Intuition.

Kahneman, D. (2011) Thinking, Fast and Slow. Farrar, Straus and Giroux, New York, USA.

Kantowitz, B.H. and Sorkin, R.D. (1983) Human Factors: Understanding People—System Relationships. Wiley, New York.

Kohn, L., Corrigan, J. and Donaldson, M. (2000) To Err is Human: Building a Safer Healthcare System, p. 28. National Academy Press.

Kostopoulou, O. and Delaney, B. (2007) Confidential reporting of patient safety events in primary care: results from a multilevel classification of cognitive and system factors. Qual Saf Health Care, 16, 95–100.

Kühl, T., Eitel, A., Scheiter, K. and Gerjets, P. (2014) A call for an unbiased search for moderators in disfluency research: reply to oppenheimer and alter (2014) Appl. Cognitive Psychol.

Lavie, N. (2005) Distracted and confused? Selective attention under load. Trends Cogn. Sci., 9, 75–82.

Lavie, N. and De Fockert, J.W. (2003) Contrasting effects of sensory limits and capacity limits in visual selective attention. Percept. Psychophys., 65, 202–212.

Macdonald, J.S.P. and Lavie, N. (2011) Visual perceptual load induces inattentional deafness. Atten. Percept. Psychophys., 73, 1780–1789.

Martins, I.T., Martins, E.T., Soares, M.M. and da Silva, A.L.G. (2013) Human Error in Aviation: The Behavior of Pilots Facing the Modern Technology. In HCI (11), pp. 150–159.

Murphy, S., Fraenkel, N. and Dalton, P. (2013) Perceptual load does not modulate auditory distractor processing. Cognition, 129, 345–355.

Obradovich, J. and Woods, D. (1996) Users as designers: how people cope with poor HCI design in computer-based medical devices. Hum. Factors, 38, 574–592.

Salthouse, T.A. and Saults, J.S. (1987) Multiple spans in transcription typing. J. Appl. Psychol., 72, 187.

Soboczenski, F., Cairns, P. and Cox, A.L. (2013) Increasing Accuracy by Decreasing Presentation Quality in Transcription Tasks. In Human–Computer Interaction—INTERACT 2013, pp. 380–394. Springer, Berlin, Heidelberg, New York.

Song, H. and Schwarz, N. (2008) Fluency and the detection of misleading questions: low processing fluency attenuates the Moses illusion. Soc. Cognition, 26, 791–799.

Sörqvist, P. (2014) On interpretation and task selection in studies on the effects of noise on cognitive performance. Front. Psychol., 5, 1249.

Stanovich, K.E. and West, R.F. (2002) Individual differences in reasoning: implications for the rationality debate. Behav. Brain Sci., 23, 645–726.

Tellinghuisen, D.J. and Nowak, E.J. (2003) The inability to ignore auditory distractors as a function of visual task perceptual load. Percept. Psychophys., 65, 817–828.

Thimbleby, H. (2007) Ignorance of Interaction Programming is Killing People. In ACM Interactions, pp. 52–57.

Thimbleby, H. and Cairns, P. (2010) Reducing number entry errors solving a widespread, serious problem. J. R. Soc. Interface/R. Soc., 7, 1429–1439.

Vicente, K.J., Kada-Bekhaled, K., Hillel, G., Cassano, A. and Orser, B.A. (2003) Programming errors contribute to death from patient-controlled analgesia: case report and estimate of probability. Can. J. Anaesthesia, 50, 328–332.

Wallace, D.R. and Kuhn, R.D. (2001) Failure modes in medical device software: an analysis of 15 years of recall data. Int. J. Reliab. Qual. Safety Eng., 8, 351–371.

Wiklund, M.E. (1995) Medical Device and Equipment Design: Usability Engineering and Ergonomics. Taylor & Francis. http://books.google.co.uk/books?id=fLybkzNdQSIC0.

Wiseman, S., Cairns, P. and Cox, A. (2011) A Taxonomy of Number Entry Error. HCI2011: The 25th BCS Conference on Human-Computer Interaction, p. 29.

Wiseman, S., Cox, A., Gould, S. and Furniss, D. (2012) Errordiary: Support for Teaching Human Error. Paper Presented at the Contextualised Curriculum Workshop at CHI 2012, Austin, TX.

Zhang, J., Johnson, T.R., Patel, V.L., Paige, D.L. and Kubose, T. (2003) Using usability heuristics to evaluate patient safety of medical devices. J. Biomed. Inform., 36, 23–30.